THE UNIVERSITY
of EDINBURGH

Spectrally and Temporally Resolved Single Photon Counting in Advanced Biophotonics Applications

András Kufcsák

THE UNIVERSITY *of* EDINBURGH

A thesis submitted for the degree of Doctor of Philosophy

2019

Borókának és szüleimnek

Abstract

Biomedicine requires highly sensitive and efficient light sensors to analyse light-tissue or light-sample interactions. Single-photon avalanche diode (SPAD) sensors implemented with complementary metal-oxide-semiconductor (CMOS) technology have a growing range of applications in this field. Single-photon detection coupled with integrated timing circuits enables us to timestamp each detected photon with high temporal resolution (down to picoseconds). Arrays of SPAD based pixels and CMOS technology offer massively parallel time-resolved single-photon counting for spectrally and temporally resolved analysis of various light phenomena.

This thesis examines how time-resolved CMOS SPAD based line sensors with per pixel timing circuits can be utilized to advance biophotonic applications. The study focuses on improving the existing techniques of fluorescence and Raman spectroscopy, and demonstrates for the first time CMOS SPAD based detection in optical coherence tomography (OCT). A novel detection scheme is proposed combining low-coherence interferometry and time-resolved photon counting. In this approach the interferometric information is revealed from spectral intensity measurements, which is supplemented by time-stamping of the photons building up the spectra.

Two CMOS SPAD line sensors (Ra-I and its improved version, Ra-II) were characterized and the effect of their parameters on the selected techniques was analysed. The thesis demonstrates the deployment of the Ra-I line sensor in time-resolved fluorescence spectroscopy with indications of the applicability in time-resolved Raman spectroscopy. The work includes integration of the sensor with surrounding electrical and optical systems, and the implementation of firmware and software for controlling the optical setup. As a result, a versatile platform is demonstrated capable of micro- and millisecond sampling of spectral fluorescence lifetime changes in a single transient of fast chemical reactions.

OCT operating in the spectral domain traditionally uses CMOS photodiode and charge-coupled device (CCD) based detectors. The applicability of CMOS SPAD sensors is investigated for the first time with focus on the main limitations and related challenges. Finally, a new detection method is proposed relying on both the wave and particle nature of light, recording time-resolved interferometric spectra from a Michelson interferometer. This method offers an alternative approach to analyse luminous effects and improves techniques based on the light's time of flight. As an example, a proof of concept study is presented for the removal of unwanted reflections from along the sample and the optical path in an OCT setup.

Acknowledgements

This work was funded by the Engineering and Physical Sciences Research Council (EPSRC, UK) (EP/K03197X/1)

Firstly, I am thankful for the enormous help received from my main supervisor at the University of Edinburgh, Dr. Nikola Krstajić. His expertise and guidance was invaluable for achieving the results presented in this work. Without his support, encouragement and patience, this thesis could not have been accomplished. I would also like to thank Professor Robert Henderson and Professor Kev Dhaliwal, my second and third supervisor for their support. Their knowledge and insight into SPAD technology and biology is a source this thesis greatly benefited from.

Photonics in life sciences demands expertise from many disciplines, where breakthroughs can only happen as a result of joint team effort and active collaboration. Throughout the duration of this study I had the privilege to work with a great number of highly skilled researchers specialized in different branches of biology, chemistry, medicine, physics and engineering. While it would not be possible to mention each person who contributed to the work described in this thesis one way or another, I would like to highlight and thank for the help and time of a few individuals. I am grateful for Philip Emanuel, Dr. Emma Scholefield and Dr. Alicia Megia-Fernandez for their contribution with sample preparation for experiments in time-resolved fluorescence spectroscopy. I would also like to thank for the assistance I received from Dr. Pierre Bagnaninchi, who provided fundamental support in the studies in optical coherence tomography. A significant help was received from constructive discussions of optics and physics with Katjana Ehrlich, Dr. Mike Tanner, Drs. Gareth and Elvira Williams, for which I would like to express my gratitude. Engineers of the CMOS Sensors and Systems group must be equally acknowledged. Dr. Richard Walker, Dr. Ahmet Erdogan and Tarek Al Abbas in particular provided an endless source of knowledge revealing the secrets of SPADs for me, for which I am most grateful.

I would like to thank EPSRC and Medical Research Council (MRC) Centre for Doctoral Training in Optical Medical Imaging, OPTIMA, (EP/L016559/1) for access to a supercontinuum laser which is an elemental component of the applications presented in this work.

Beyond all the academic contribution, I received the greatest support from my partner, Boróka Vas, and my family. Their endless love and encouragement cannot be overestimated. I will always be grateful for them for standing by me throughout this journey.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

András Kufcsák

Contents

List of Figures

List of Tables

Acronyms

ADC analogue-to-digital converter
APD avalanche photodiode
API application programming interface
APS active pixel sensor

BSI backside illumination

CARS coherent anti-Stokes Raman spectroscopy
CCD charge-coupled device
CMM 'centre of mass' mode
CMOS complementary metal-oxide-semiconductor
CPS counts per second

DAC digital-to-analogue converter
DCM digital clock manager
DCR dark count rate
DDG digital delay generator
DNL differential nonlinearity
DOF depth of focus
DR dynamic range
DSP digital signal processor

EMCCD electron multiplying charge-coupled device

FAM 5-carboxyfluorescein
FCS fluorescence correlation spectroscopy
FD frequency-domain
FD OCT Fourier-domain optical coherence tomography
FDML Fourier-domain mode locking
FF fill factor
FFT fast Fourier transform
FIFO first in first out
FLIM fluorescence lifetime imaging microscopy
FOV field of view
FPGA field-programmable gate array
FRC full-range complex
FRET Förster resonance energy transfer

FSI frontside illumination
FSM finite-state machine
FW firmware
FWHM full width at half maximum

GPGPU general-purpose graphical processing unit
GUI graphical user interface

ICCD intensified charge-coupled device
IIR infinite impulse response
INL integral nonlinearity
IP intellectual property
IR infrared
IRF instrument response function

KID kinetic inductance detector

LED light-emitting diode
LIDAR light detection and ranging
LRRS low-resolution Raman spectroscopy
LS least squares
LSB least significant bit
LUT lookup table

MCP microchannel plate
MLE maximum likelihood estimation
MMP matrix metalloproteinase
MOS metal-oxide-semiconductor
MOSFET metal-oxide-semiconductor field-effect transistor
MR methyl red

NA numerical aperture
ND neutral density
NIM nuclear instrument module
NIR near infrared

OCT optical coherence tomography
OK Opal Kelly

PCB printed circuit board
PCI Peripheral Component Interconnect
PDE photon detection efficiency

PDP photon detection probability
PLL phase-locked loop
PMT photomultiplier tube
PSD power spectral density
PSF point spread function

QE quantum efficiency
QIS quanta image sensor

RLD rapid lifetime determination
RRS resonance Raman spectroscopy

SC supercontinuum
SD spectral-domain
SD OCT spectral-domain optical coherence tomography
SDK software development kit
SERDS shifted-excitation Raman difference spectroscopy
SERS surface-enhanced Raman spectroscopy
SLD superluminescent diode
SMLM single molecule localisation microscopy
SNR signal-to-noise ratio
SNSPD superconducting nanowire single-photon detector
SOC system on chip
SPAD single-photon avalanche diode
SPC single-photon counting
SRS stimulated Raman spectroscopy
SS OCT swept-source optical coherence tomography
SSPD superconducting single-photon detector
STI shallow trench isolation
STJ superconducting tunnel junction

TAC time-to-analogue converter
TCSPC time-correlated single-photon counting
TD time-domain
TD OCT time-domain optical coherence tomography
TDC time-to-digital converter
TERS tip-enhanced Raman spectroscopy
TES transition-edge sensor
TOF time-of-flight
TRFS time-resolved fluorescence spectroscopy

TTL transistor-transistor logic

VECSEL vertical-cavity surface emitting laser
VPH volume phase holographic

Chapter 1

Introduction

This thesis focuses on optical spectroscopy instrumentation and in particular, the timing aspect behind optical detection. This thesis studies a number of biomedical optics applications in the time-resolved domain, as explained in this chapter and chapter 2. As described below, optical spectroscopy has a long history, but the nanosecond and sub-nanosecond events have only been studied in the last 50–60 years. Furthermore, integrating a massive number of time-resolved detectors with resolutions approaching 10 picoseconds has become possible with innovations in electronics. The challenge now is to exploit these new detectors in existing techniques, but also, in perspective, to invent completely new techniques altogether. This chapter introduces the reader to optical spectroscopy and its history, while giving a brief overview of existing detectors. The structure of the thesis is also given with enough material to tackle time-resolved aspects in chapters 2 and 3.

1.1 Spectroscopy

The ability to sense electromagnetic radiation was a major step in evolution dating back to millions of years ago (Feuda *et al.*, 2012). Vision is the main source of information to humans and for us who have good vision, we are continuously amazed by the colours and shapes surrounding us (Haupt and Huber, 2008). Therefore, understanding what light is and furthermore, describing its properties (e.g. colours) has challenged us since the dawn of humanity. It is only recently that we have had some success in designing scientific instruments to study the small and the big using light (e.g. microscope was discovered in last 400–500 years). Performing medical diagnosis using light is a step further, and could only follow the discoveries in what we now call physics (Natural Philosophy in Newton's time). Spectroscopy is one these techniques where enormous amount of information can be gleaned by splitting light into constituent colours using devices such as prisms and diffraction gratings (Fig. 1.1).

Traditionally, the colours of the incident light are separated using a prism. In such medium the velocity that each light component travels with (phase velocity) depends

on the component's frequency. A consequence of this is that the refractive index of the components is also different. Spatial separation occurs as each component of different frequency (different colour) has a different angle of refraction when entering and exiting the prism. The range of colours from light shining through a piece of glass was already noted in ancient Roman times (Hine, 2006). Still, for centuries it was thought that colours originate from the material alone. In his famous scientific study from 1671, Isaac Newton showed that colours are an inherent property of light, white light can be split up to a range of colours and rays of light with different colours can be recombined again to white light (Newton, 1671). Newton named the range of colours appearing from white light after the Latin 'spectrum', meaning image, apparition. Newton's explanations trying to make sense of his results reinforced the corpuscular theory, explaining light as being formed of discrete particles. Even though his theory does not provide a complete picture, Newton's work was an important milestone towards understanding the nature of light.

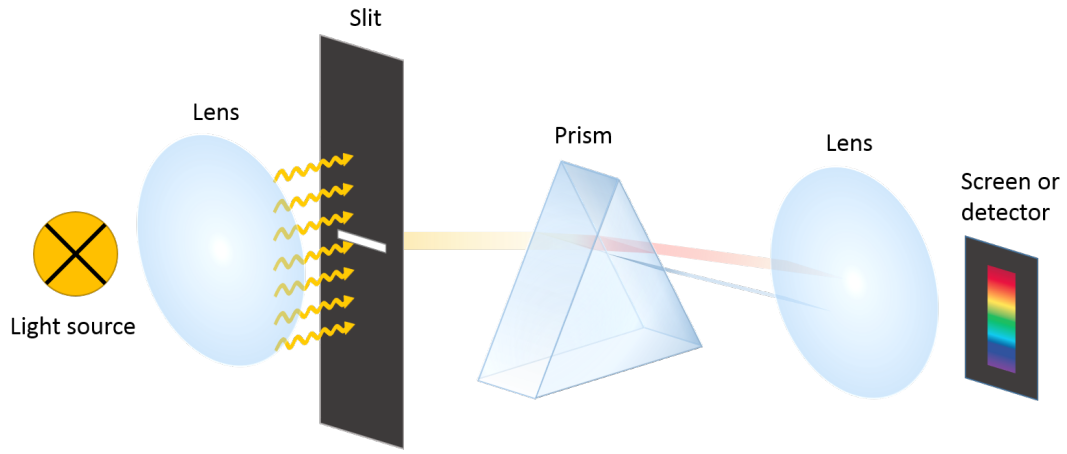

Figure 1.1: Basic optical configuration of a spectrometer.

Spectroscopy was further developed throughout the 19th century (Burns, 1987, 1988; Thomas, 1991; Hearnshaw, 2014) by adding a second lens to the instrument (Simms, 1840), and using a grating instead of prism which disperses light using diffraction (Fraunhofer, 1821). These instruments had a much higher resolving capability and higher accuracy. At the beginning of the century, it was also realized that dispersed white light has more components than those of which the eyes could see. In 1800 William Herschel identified non-visible components beyond red colour, the infra-red radiation (Herschel, 1800). In 1801 J. W. Ritter recognized components over the purple end of the spectrum, ultra-violet radiation (Ritter, 1801). Analysing light emitted from luminous materials such as flames of salts (Herschel, 1823; Talbot, 1826), sparks of metals (Wheatstone, 2011), and hot gases led to the realization that bright lines of the emitted spectra (Fig. 1.2) are characteristic to the elements under investigation.

Similarly, when studying the spectra of distant stars (Huggins and Miller, 1864), solar light (Wollaston, 1802; Kirchhoff, 1860) or light absorbed by cold gases, the resulted spectra contain dark lines which are also specific to the chemical constituents of the subject (Fig. 1.2). Spectroscopy soon become a fundamental tool of analytical chemistry, as it become clear that the emission and absorption spectra are signatures of chemical elements. The success of absorption and emission spectroscopy led to the discovery of several previously unknown elements (Weeks, 1932).

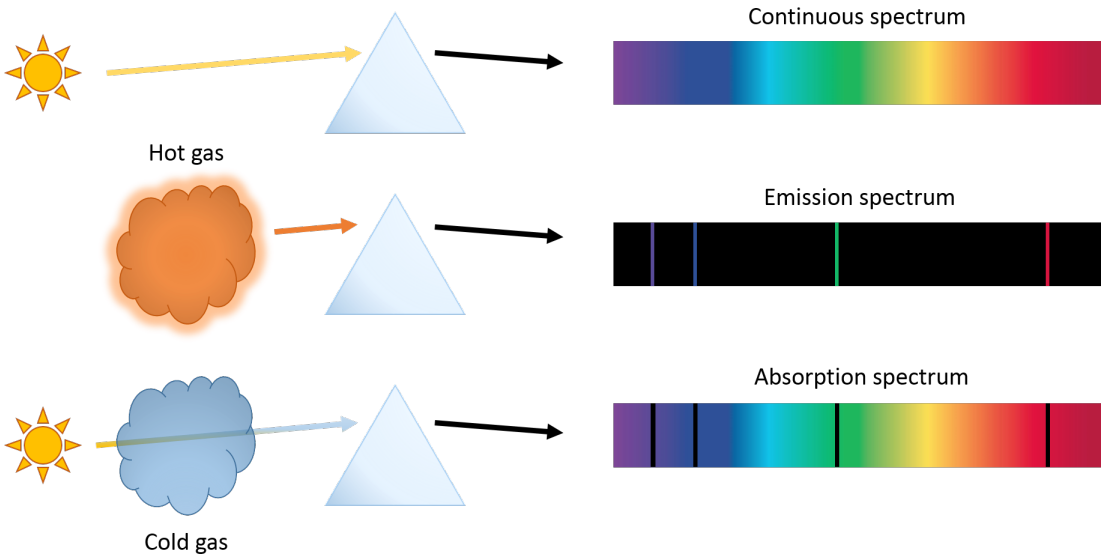

Figure 1.2: Optical spectroscopy as a tool to study emission from objects and gases and how sharp emission peaks and sharp absorption lines increased our understanding of the nature of light and matter.

Even though spectroscopy had been widely used in physics, chemistry and astronomy, no concise explanation existed how the emission and absorption lines are formed. Finally, in the 1920s the discoveries in quantum mechanics explained the phenomenon by atomic orbitals and fixed energy levels of the electrons (Lakowicz, 2006, chapter 1 and 2). According to this interpretation, electrons can be promoted to a higher energy level by absorbing a photon with the energy matching the difference of the two levels. The dark lines in the absorption spectrum are characteristic to the electron structure of the atoms absorbing the light (Fig. 1.3a)). Also, when an excited electron moves to an available level of lower energy, it emits a photon exactly with the surplus energy. The emission spectrum is in turn characteristic to the electron structure of atoms emitting the light (Fig. 1.3b)).

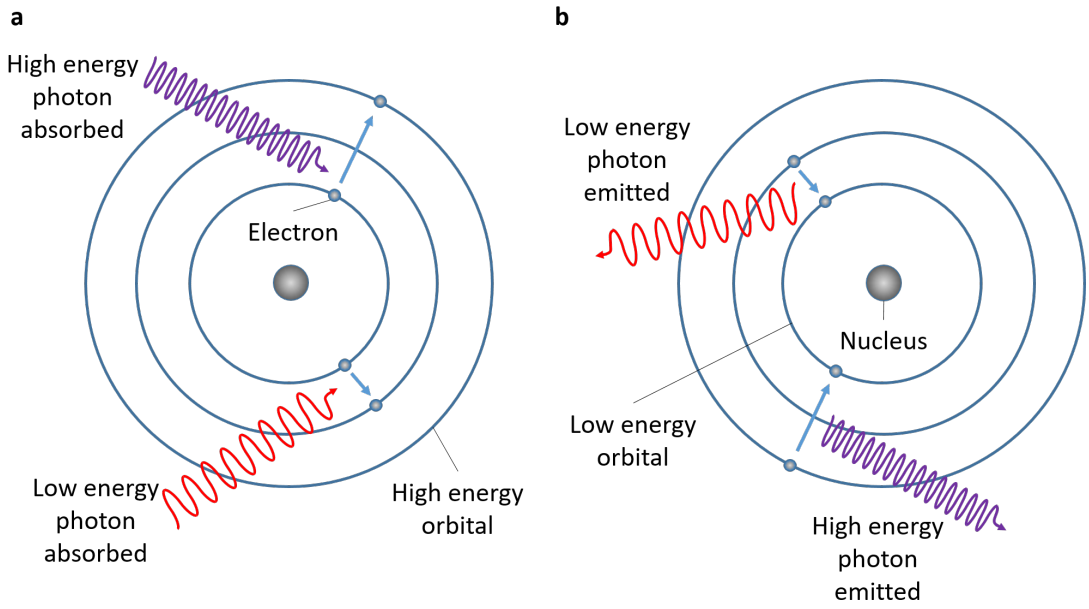

Figure 1.3: Electron structure and absorption and emission of light.

1.2 Sensors

Initially, the light from spectrometers was projected onto a screen (or observed directly by eye), which does not allow accurate, quantitative studies of minute changes and non-visible components of light. Modern spectroscopy would not exist without light detection devices converting the optical information to electric signals. Most light sensors work on the same principle: they convert the incident optical power into the generation of electric charge carriers. Early demonstrations revealed that light can generate voltage and current (photovoltaic effect) (Becquerel, 1839) or change the conductivity of materials (photoconductivity) (Smith, 1873). Light sensors based on these phenomena such as photoresistors (light-dependent resistors) and photocells have limited accuracy and low sensitivity, however they are still used nowadays in applications with low requirements.

Extensive studies on charged bodies (Hertz, 1887; Thomson, 1906) also showed that certain materials emit electrons upon the absorption of light (photoelectric effect). Classical electromagnetic theory, where light is treated as an electromagnetic wave, could not explain this phenomenon. According to classical model, the intensity of incident light would change the kinetic energy of the emitted electrons. Also, low intensity of the incident light should cause a time lag between light excitation and electron emission. Instead, the emission of electrons happens instantaneously, but only above a characteristic frequency threshold of the incident light. Increasing the intensity of light increases the number of emitted electrons. The contradiction between

theory and observations was resolved by Albert Einstein, who (in agreement with Max Planck's findings on black-body radiation) gave explanation to the photoelectric effect (Einstein, 1905). He suggested that light is composed of discrete quanta of energy. Light quanta later were given the name 'photons' from the Greek word for light (Kragh, 2014). Einstein's explanation of the photoelectric effect, for which he received the Nobel price in 1921, was a milestone towards the concept of wave-partical duality.

Low light levels with minute electrical effects are hard to detect. Therefore most light detection devices utilize some kind of amplification mechanism. One example with widespread applications is the photomultiplier tube (PMT), invented in 1934 (Iams and Salzberg, 1935; Zworykin *et al.*, 1936; Kubetsky, 1937). In PMTs, light induced photo-electrons are multiplied through secondary emission (Fig. 1.4). In this effect, the free photo-electrons (primary electrons) are accelerated towards an electrode of high electrical potential (called a dynode). The high kinetic energy free electrons generate further electron emission upon impact with the dynode. Several dynodes with increasingly higher electrical potential provide an exponential multiplication of electrons. The achievable gain is between 10^6 - 10^8, turning the original photo-electrons to an easily detectable, sharp current pulse.

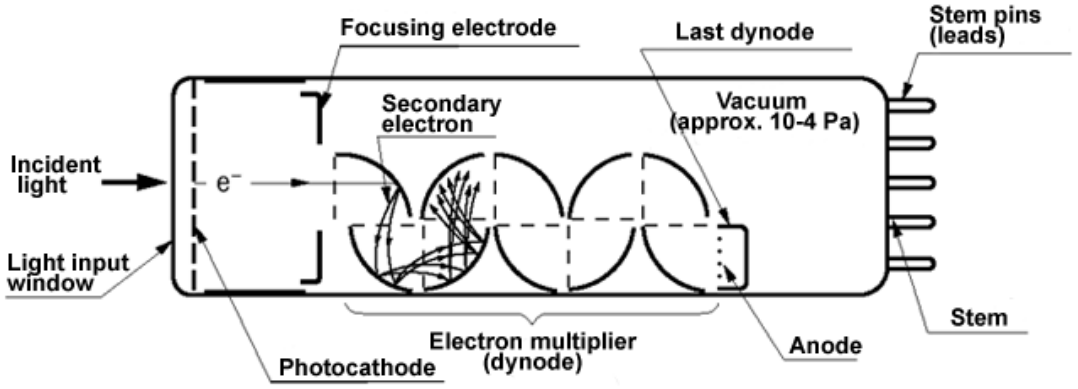

Figure 1.4: Basic structure of a photomultiplier tube. Figure from Hamamatsu Photonics (Hamamatsu, 2007).

PMTs have been employed in a myriad of applications across the fields of microscopy, spectroscopy, astronomy and medical diagnostics to name only a few. Their advantages include high gain, high sensitivity and low noise. The main drawbacks of PMTs are their size and bulky construction. Further weaknesses involve the necessity for additional hardware of the high voltage supply and processing electronics, and PMTs' susceptibility to magnetic fields. The size of photomultiplier tubes is much reduced in microchannel plate (MCP) detectors, where glass capillaries of 6–20 µm diameter are bundled together in parallel, with a spacing of ~15 µm. The capillaries

of the array form electron multiplier channels with a continuous dynode (Fig. 1.5). These devices still require a high voltage across the channels, furthermore they have a high gain variance and short lifetime. New technologies, with the emphasis on solid-state integrated circuits have been gradually replacing PMTs and MCPs. A detailed discussion of PMTs is given by Becker (2005, chapter 6).

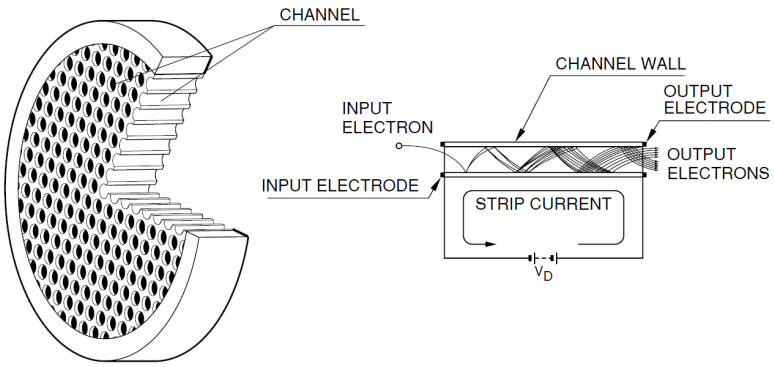

Figure 1.5: Basic structure of a microchannel plate. Figure from Hamamatsu Photonics (Hamamatsu, 2007).

1.2.1 Semiconductor sensors

The rise of semiconductor electronics quickly revolutionized the technology for light sensing. Incident light generates electric charge in the semiconductor crystal structure. Solid state devices and circuits are convenient for the accumulation, transfer and amplification of electrons and holes generated in light sensitive semiconductor structures (e.g. PN junctions of photodiodes). In comparison with PMTs, light sensing semiconductor structures are all-round devices with compact size allowing large array formats and highly integrated system on chip (SOC) solutions.

In charge-coupled devices (CCDs) the light sensitive structures of pixels are metal-oxide-semiconductor (MOS) capacitors (Fig. 1.6). Light generated free electrons in a p-type MOS capacitor are collected and stored in a potential well, controlled by a positively biased gate electrode. The collected charge on the pixel, which is proportional to the light exposure, can be transferred to the neighbouring well by controlling the voltages applied to the gate structure. In its simplest form, the pixel charges are shifted in subsequent stages along parallel lines of pixels of a two dimensional array during readout (parallel shift in Fig. 1.9). At the end of the parallel pixel rows, a line of charges is shifted and measured, one at a time (serial shift in Fig. 1.9). This step includes amplifying and turning the charges to voltages, followed by digitizing and further processing. Slightly different structures and readout methods exist allowing higher frame rates.

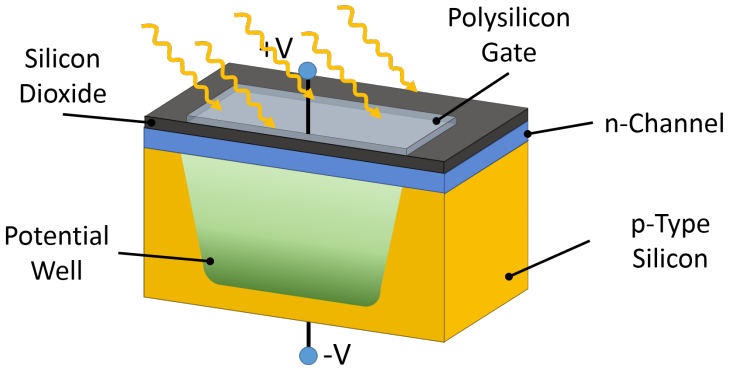

Figure 1.6: Basic structure of a CCD pixel.

Due to their high sensitivity, compact size, high fill factor (ratio of light sensitive area to the whole surface) and quantum efficiency (capability to convert photons into charge carriers), CCD based sensors formed the predominant technology for light detection in scientific instruments and consumer electronics for decades. Acknowledging the excellence of these sensors, George E. Smith and Willard Boyle, inventors the CCD technology received the Nobel price in physics in 2009. The main drawback of CCD sensors is the imperfect charge transfer, restraining the number of pixels and readout rates. CCDs are furthermore hard to integrate with digital logic, hindering their applicability in increasingly complex applications as miniaturized smart sensors.

To overcome the limitations of CCD sensors introduced by imperfect charge shifting, sensors with active pixel design started to emerge in around the 1990's (Fossum, 1993). In an active pixel sensor (APS), each photodetector element comprises one or more transistors for buffering and amplifying the signal, which is driven off the pixel using metallic wires. At the beginning, APSs were inferior to CCDs in many respects, as the additional transistors introduced more complexity and reduced the photosensitive area (resulting in lower fill factor and light efficiency). APSs started to be competitive with CCDs with advancements in photolitography allowing sub-micrometer transistor sizes. The increasing dominance of complementary metal-oxide-semiconductor (CMOS) technology for manufacturing integrated circuits, and applying microlens arrays over the APS pixel matrix to increase the fill factor led to improved designs and wider adoption. Typical CMOS design uses pairs of n-type and p-type metal-oxide-semiconductor field-effect transistors (MOSFETs) leading to much lower power consumption. Less specialized manufacturing needs and lower power consumption facilitated CMOS APSs to take over CCDs, especially in consumer electronics.

In most APSs, the photodetector element is a reverse biased photodiode (Fig. 1.7), with the bias far below the breakdown voltage of the diode. In photodiodes, the energy of incident photons create electron-hole charge carrier pairs, similarly to CCDs. If

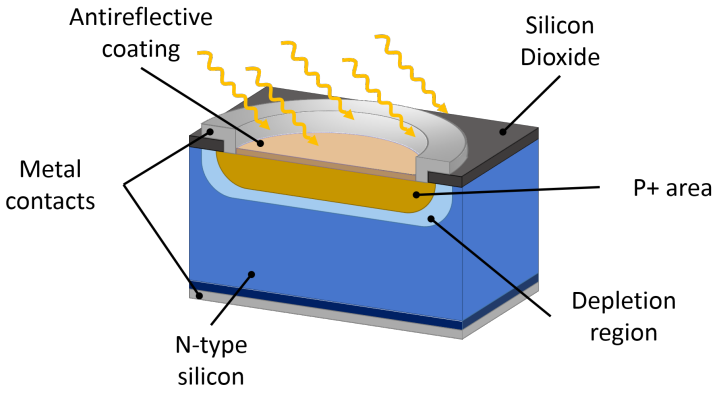

Figure 1.7: Basic structure of a photodiode.

these charge carriers are generated in the depletion region of the PN junction of the diode (or close to it so they can reach the depletion region with diffusion), they are swept from the junction by the electric field. Due to the applied bias, free electrons move further to the positive terminal of the diode while holes move to the negative terminal, producing a current. The reverse bias furthermore increases the width of the depletion region which in turn increases the number of photo-generated charge carriers contributing to the current. With an increasing depletion region width, the capacitance of the PN junction decreases which reduces the response time of the photodiode. This basic design can be further improved by having a wide, undoped intrinsic region between the P and N-type layers of the diode (PIN diode). The intrinsic region further widens the depletion region and hence increases the sensitivity of the diode through increasing quantum efficiency, and also reduces its response time by decreasing the diffusion time of charge carriers generated close to but outside the depletion region.

1.2.2 Single-photon sensitivity

With sufficient amplification and low detection noise, it is possible to measure a signal induced by a single photoelectron and reach the quantum limit of light detection, sensing a single photon. Single-photon counting (SPC) is a technique where light levels are studied as an aggregation of digital events. This technique has been widely applied in low light level scenarios in biology, chemistry, such as single-photon microscopy, Raman spectroscopy, or fluorescence analysis (Becker, 2005, chapter 6; Lakowicz, 2006, chapter 4). Detecting photon arrivals as digital events also makes it possible to study the time distribution of optical signals using high resolution timing electronics (see next section). Photon counting light detectors do not necessarily record every photon coming from a source, due to several properties of the detector such as the photon absorbing material or the detector's working principle. The sensors' material

determines which wavelengths of photons can be absorbed and the probability of charge carrier generation (quantum efficiency). Another effect preventing the detector to capture photons is temporal insensitivity after the detection of a previous photon event, known as deadtime. Undetected photons do not represent a problem in a photon counting application if the measured photon counts are coming at a rate lower than the maximum count rate (inverse of deadtime), so that the changes in photon counts reflect the changes in the observed light levels. The output of the detectors is also affected by false events which do not correspond to real photons at the input, representing noise. Sources of this noise could be thermally generated charge carriers, noise of the detector electronics, repeated signalling of a photon event, known as afterpulsing, or crosstalk between sensing elements.

Photomultiplier tubes and microchannel plates

PMTs and MCPs with high gain and low noise have been applied for photon counting. The amplified output signal of the device is used to mark a photon absorption either by comparing it to a preset threshold level, or when it reaches a certain fraction of the total peak height. Photon counting mode operation of PMTs and MCPs have several advantages over an analogue light detection mode, such as increased stability over varying gain of the PMT, and increased signal-to-noise ratio of detection (Hamamatsu, 2007, chapter 6). PMTs and MCPs are still the most popular single point sensors in confocal and multiphoton fluorescence microscopy, due to their low noise ($<$1 Hz dark counts) and low uncertainty of detection time (~20 ps). However, as mentioned above, their size and cost limit widespread deployment, while setups running on single detector are slow in imaging applications.

Charge-coupled devices

In the case of CCD sensors, single-photon sensitivity has been achieved with intensified charge-coupled devices (ICCDs) and electron multiplying charge-coupled devices (EMCCDs) (Hynecek, 2001). The main manufacturers of these scientific grade CCDs include Andor, UK and Princeton Intruments, USA. ICCDs are CCD based detector systems where an additional intensifier detects and amplifies an image which is then projected onto the CCD with coupling lenses or fibre optics. The intensifier consists of a photocathode generating free electrons upon photon absorption, an MCP for multiplying the electrons through secondary emission, and a phosphor layer emitting photons upon electron impingement (Fig. 1.8). Using an intensifier provides much higher sensitivity and allows shuttering of the light signals (see next section). The intensifier, however, also introduces a deterioration of spatial resolution, due to intermediate optics. Furthermore, high gains negatively affect the achievable dynamic range (defined in Chapter 3).

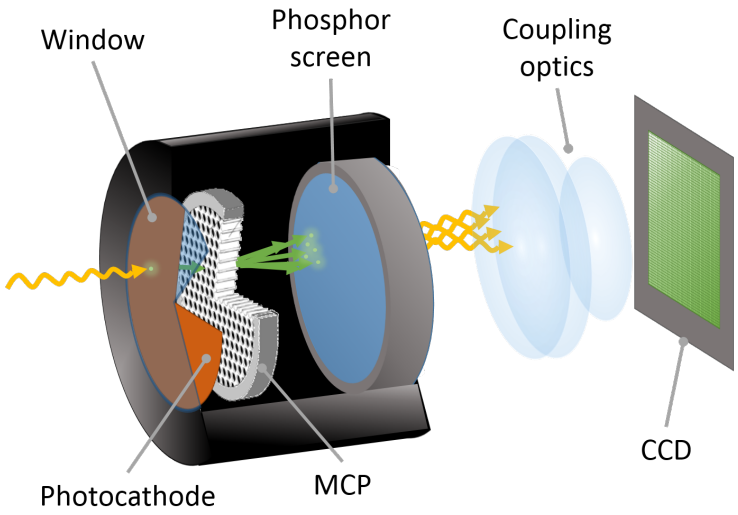

Figure 1.8: Basic structure of an intensified CCD sensor. The image of the MCP is from Hamamatsu Photonics (Hamamatsu, 2007).

CCDs based sensors deploy an analogue-to-digital converter (ADC) to sample the incoming signal. For each readout ADCs reproduce the signal with quantization and readout noise. While quantization noise can be reduced by using a 16 bit ADC as opposed to 8 bit ADC, the readout noise is an inherent property of the ADC that compromises single-photon sensitivity. Electron multiplying CCDs detect light as regular CCDs, conserving most of the parameters of a CCD but overcoming the limitations of the readout noise. The difference with EMCCD technology is in the last step before measuring the detected charge per pixel, where a line of charges is stepped through an electron multiplying register (Fig. 1.9). The EM register is built of several hundred cells. When stepped to a neighbouring cell of higher voltage, charges acquire energy which is turned to the generation of an electron-hole pair with small probability. Via these 'impact ionization' events, the charges are amplified as much as thousand times when reaching the readout electronics. Unfortunately, this process also amplifies thermally generated electrons as well, increasing the noise level. To prevent this, cooling is crucial for EMCCDs, which increases the cost of the device and could lead to problems due to condensation. As a result of the amplification of the optical signals, the readout noise in EMCCDs has a lower significance compared to CCDs. However, the stochastic nature of electron multiplication introduces uncertainties in the number of generated electrons. This effect is expressed as a multiplicative term increasing the noise from statistical uncertainties of the generated photoelectrons (shot noise) and dark counts, often denoted as excess noise factor. In contrast with single channel PMTs, EMCCDs are 2D devices acquiring images in one go, although their conversion time is not as low and reliable as in PMTs and MCPs (Becker, 2005, chapter 6;

Hamamatsu, 2007).

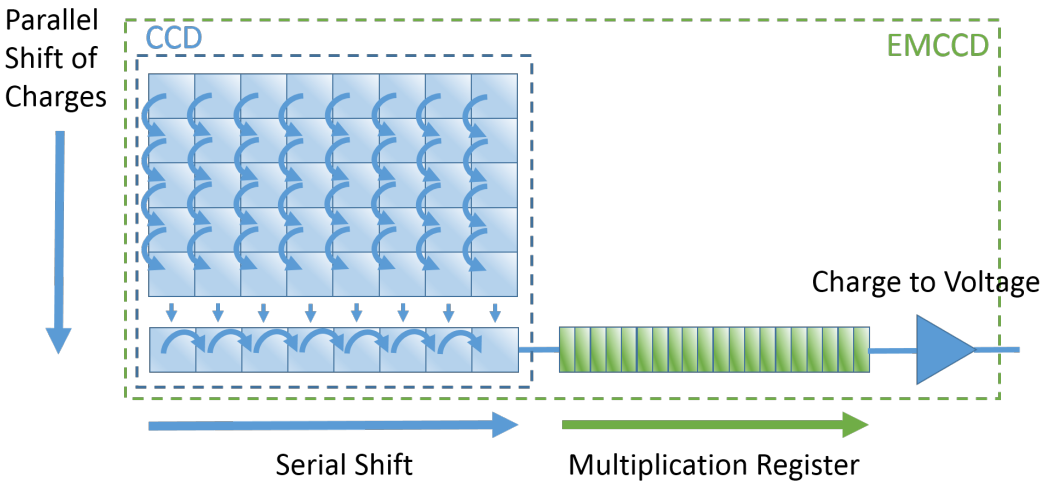

Figure 1.9: Basic structure of an EMCCD sensor.

Silicon photodiodes

Increased sensitivity can be achieved with silicon photodiodes as well through multiplication of charge carriers. In avalanche photodiodes (APDs) where the reverse bias is close to the breakdown voltage, the high electric field across the device accelerates electrons and holes, causing them to generate more and more charge carriers through impact ionization. The resulting avalanche of charge carriers is proportional to the incident light activity and easy to detect. These devices have excellent timing characteristics due to the fast buildup of the avalanche current. The avalanche process amplifies the noise (thermally generated carriers) as well. With additional amplifiers, avalanche currents caused by single photons could be sensed, providing single-photon sensitivity of the APD. The additional amplification, however, increases the timing uncertainty of photon detection and noise of the APD.

When the bias voltage is above the APD's breakdown voltage, the APD stays in an unstable state for a limited amount of time and does not conduct current. When a photon is absorbed during this time, the generated avalanche current has virtually an infinite gain, and the output of the APD can be sensed easily without further amplification (Fig. 1.10). The presence of the output signal is regarded as a digital event signalling the detection of a single photon. This mode of operation is called Geiger-mode, and APDs used in this mode are called single-photon avalanche diodes (SPADs). The strong avalanche current can damage the SPAD and hence needs to be quenched. Passive quenching mechanisms use a ballast resistor in series with the diode to reduce the bias voltage over the SPAD upon an avalanche breakdown.

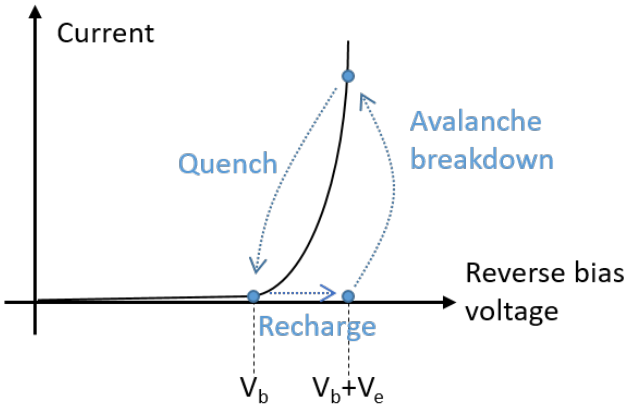

Figure 1.10: Basic operation of single-photon avalanche diodes (V_b - breakdown voltage, V_e - excess bias voltage).

Active quenching circuits can reduce the bias voltage upon sensing an avalanche breakdown, and recharge the device thereafter. The result is a shorter reset time than passive solutions at the price of increased complexity. High sensitivity and fast timing capabilities of SPADs in big arrays come at trade-offs between several parameters of the sensor. CMOS SPADs are the main enabling technology in this thesis and more detailed description of SPAD parameters is given in Chapter 3, where two sensors Ra-I and Ra-II are introduced and evaluated.

Superconducting detectors

In superconducting detectors the absorption of photons is generally sensed by detecting changes in conductivity. Superconducting detectors are particularly attractive to capture photons of infrared (IR) wavelengths, where most semiconductor devices are restricted in efficiency. Examples of single-photon sensitive, superconducting detectors include superconducting tunnel junctions (STJs) (Peacock *et al.*, 1996), transition-edge sensors (TESs) (Irwin and Hilton, 2005), kinetic inductance detectors (KIDs) (Day *et al.*, 2003), and superconducting nanowire single-photon detectors (SNSPDs) (Gol'tsman *et al.*, 2001), which offer the highest performance of all.

An SNSPD is a wire made of a sheet of superconducting material (traditionally niobium nitride - NbN) of few nm thickness and few hundred nm width. The wire is cooled down below its critical temperature to a few K and biased with a current just below its critical current (Fig. 1.11i). When a photon of sufficient energy is absorbed in the material it causes a nonequilibrium perturbation creating a 'hot spot' and a local temperature above the critical temperature (Fig. 1.11)ii). As the resistive hot spot increases in size it expels the bias current to the sidewalks of the wire (Fig. 1.11iii). When the increased density of the supercurrent exceeds the critical threshold, the

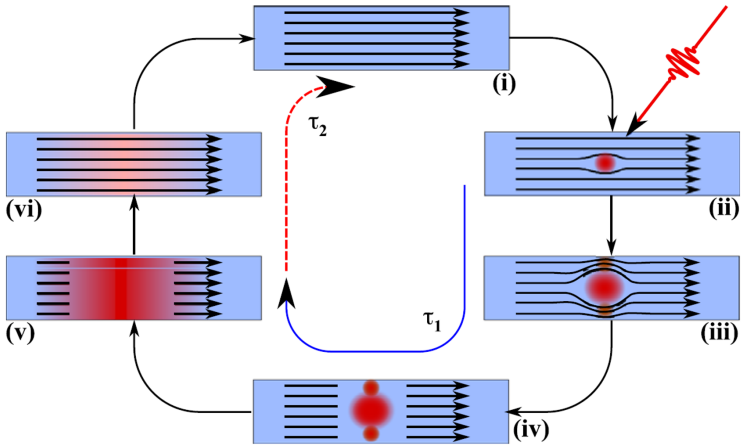

Figure 1.11: Detection cycle of a superconducting nanowire single-photon detector. Figure from Natarajan *et al.* (2012).

device becomes entirely resistive (Fig. 1.11 iv and v), forcing the bias current to a shunt circuit of lower resistance where a voltage pulse is measured. At the same time, as the current density in the device decreases, the resistive barrier cools off and the detector returns to the initial superconducting state (Fig. 1.11 vi). SNSPDs offer high detection efficiency (>90 %), sub hertz dark count rate, low timing uncertainty (~150 ps), and short deadtime (~40 ns) (Marsili *et al.*, 2013). For the definition of these measures refer to Chapter 3. The main drawback of superconducting sensors is the need for extreme cooling making the detector costly and raising mobility issues.

Quanta image sensors

A reasonably new concept in light detection is represented by quanta image sensors (QISs) (Fossum, 2011; Ma *et al.*, 2017). Instead of collecting and integrating the charges in the sensing elements such as in CCD and APS designs, QIS devices aim to detect only one (or few) photons at each of its nanoscale sensing elements, or 'jots', with high quantum efficiency. In a QIS concept, millions of single photon sensitive jots are read out at high frame rates, providing single-bit planes (where each jot is either 'on' or 'off'). Spatial and temporal aggregation of bit planes promote a wide range of digital image processing methods, such as object tracking, motion compensation, wavefront correction, more complex high dynamic range imaging methods and arbitrary spatio-temporal convolution of QIS bits.

1.3 Time-resolved light detection

Most light detection methods rely on measuring intensities and the only timing element relates to speed of detection. Time-resolved detection consists of high speed detectors coupled with picosecond timing electronics. The timing electronics is key and one only needs to observe the computers and laptops around to realize the unit of time for a 4 GHz processor is 250 ps. However, techniques for ps timing have existed for long time. Since the early 90s (20th century) efforts by companies such as Becker and Hickl Gmbh introduced PC plug-in boards (ISA and PCI) with single channel timing capability down to ps, costing between £10,000 and £20,000. PMTs from Hamamatsu were biased for single-photon counting and wired to timing circuitry. The count rate was ~1 million counts per second, but in practice the real samples would be expected to be weaker in intensity, since maximum count rate would saturate the detection. Other companies joined in, PicoQuant Gmbh, and currently we have a rich ecosystem of time-resolved detectors and timing circuitry with easy-to-use software and off-the-shelf performance to dive into studying fast and even ultra-fast phenomena in physics and chemistry.

It should be mentioned that time-resolved detection can be performed in several ways and we will discuss two: time-domain and frequency modulated. Time-domain methods usually rely on single-photon detection and timing circuitry, however other methods also exist. ICCDs mentioned above also work in time domain, but here the intensifier is switched on and off in 100s of ps allowing for fast detection of phenomena across a 2D CCD sensor. Therefore, the timing is provided by intensifier switching (or time-gating) rather than timing electronics. The temporal changes of light intensity are revealed through applying a series of different time-windows with the intensifier. Streak cameras record the temporal evolution of light by transforming the temporal intensity variations to a spatial intensity profile. This is achieved by deflecting photoelectrons from the device's photocathode as they pass a pair of electrodes with a time-dependent voltage applied. The swept supply voltage causes a time-dependent deflection and different angles of the photoelectrons entering the camera's detector (typically MCPs and CCDs). The sub-picosecond timing of streak cameras is exceptional, but the number of channels (e.g. for spectral detection) is limited, and they have a considerable cost and size.

Throughout this thesis we focus on time-domain methods and time-correlated single-photon counting (TCSPC) in particular (see chapter 2 for more details on TCSPC) which requires timing electronics down to 10s of ps. A general schematic depicting the main components of our approach is shown in Fig. 1.12.

Frequency modulated methods are not covered in this thesis experimentally, but a brief mention is needed. Here, a modulated light source is used, and the emission

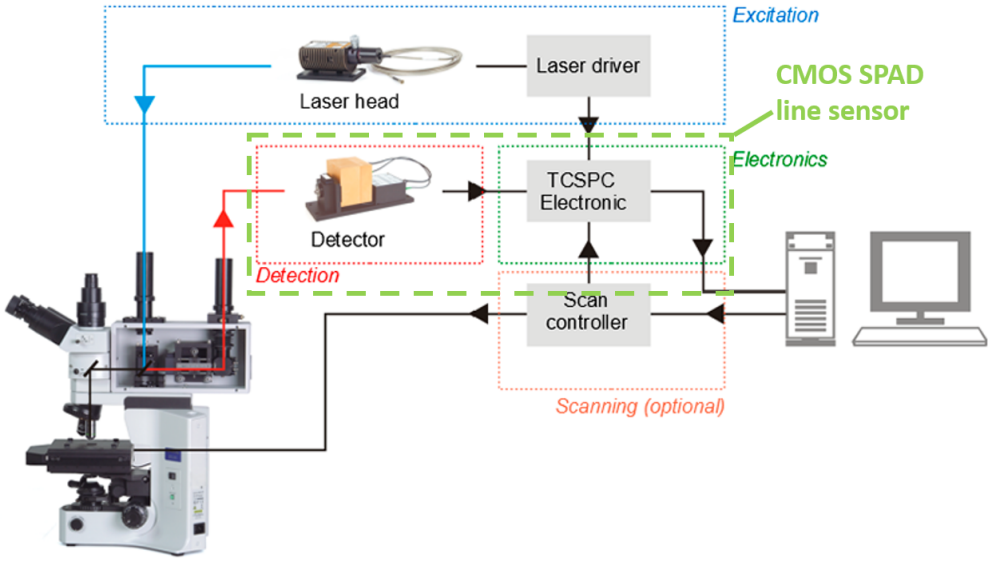

Figure 1.12: Schematic of a time-correlated single-photon counting (TCSPC) system applying a pulsed light source, a single-photon detector, TCSPC electronics and accessory optics. The key feature of CMOS SPAD line sensors is the integrated TCSPC functionality on several hundreds of channels for spectral detection. Figure is based on the image of a fluorescence microscope from PicoQuant (PicoQuant).

intensity modulation is analysed as a characteristic response to the excitation modulation frequency. The attraction of these methods is that they do not require single-photon sensitivity and allow us to use modulated sources and detectors which are readily available. In fact, it should be noted that first time-resolved studies historically were frequency modulated (early and mid 20th century) (Berezin and Achilefu, 2010; Gaviola, 1927). As an example, fluorescence emission shows a phase-lag and a change of modulation depth (peak to peak of amplitude) compared to the excitation, due to the average time spent in the excited state. From these values the apparent lifetime of the fluorescence decay can be easily calculated. Multi-exponential and more complex decays require the measurement of the frequency response to a wide range of excitation modulation frequencies. The modulation frequency needs to be around the reciprocal of the expected lifetime, typically meaning 2 to 200 MHz. Laser diodes and light-emitting diodes (LEDs) are favourable for this purpose. Observing lifetimes below ns requires GHz modulation, demanding high end light sources. Modulation of both the light sources and detectors at slightly different frequencies provides a detected signal with low frequency, but still maintaining the phase offset and change of modulation depth characteristic to the sample. The signal is digitized, and changes in phenomena and the underlying timing are inferred via Fourier transform methods. The benefit is that no perfect sinusoidal modulation is required (even a pulsed source can be applied), while the noise of the measured data is decreased. The main challenges of frequency domain methods are the direct observation of multiple lifetime compon-

ents (e.g. in video rate FLIM), and the complexity of background signal correction (Lakowicz, 2006, chapter 5).

1.4 Enabling technologies

The factors making this study possible consisted of state-of-art sensors, pulsed light sources and intermediate optics. Each of these technologies is undergoing rapid development. While this thesis focuses on exploiting latest optoelectronics circuits, one should always remember that without lasers, and pulsed lasers in particular, many of the biomedical optics techniques would not be possible.

1.4.1 Massively parallel time-resolved CMOS SPAD based sensors

Compact devices capable of massively parallel, time-resolved detection of light are enabled through SPAD based sensors manufactured with standard CMOS technology. Recent improvements in this field made it possible to produce light sensors with several hundred channels, each processing photons with high detection efficiency and tens of picoseconds timing resolution. CMOS technology has the additional benefit of allowing further functionality to be realized on the same device, which allows more complex processing of the information on the detected light levels, forming complex light detection systems on a single silicon chip. Such devices have potential of high impact with regard to spectroscopic techniques, especially in low light scenarios. The low cost of these CMOS SPAD sensors in mass production, the complex functionality and high performance while maintaining small size result in SPAD sensors penetrating other fields as well, where traditionally different technologies have been used for light detection.

1.4.2 Light sources

Lasers are by far the main enabling technology in time-resolved studies. Laser diodes in particular are affordable and intensity switching allows for pulsing down to 50 ps as sold by many suppliers (Edinburgh Instruments, Horiba, Becker and Hickl, and PicoQuant). LEDs are harder to pulse due to their switching characteristics, but 200 ps pulsed LEDs are readily available. More expensive sources include supercontinuum (SC) lasers which are broadband in emission, pulsed and can be coupled into single mode fibres, which is key for applications such as low-coherence interferometry, as will be further explained in later chapters. SC lasers are, in many ways, an ideal pulsed laser with the main disadvantage being that they are expensive (£40,000 or more with cheapest low-intensity units ~£10,000) and class 4 in terms of laser safety.

Therefore, they are mostly used in high-end microscopy and scientific instrumentation, even though they could have more widespread appeal.

1.5 Research aims

CMOS SPAD line sensors are very attractive for many low light level and photon counting applications. While their applicability is still challenging for many optical techniques, SPAD sensors have been designed with tailored detector properties to accommodate the requirements of advanced biophotonics applications. With a versatile sensor of high performance, different optical techniques could be combined in a multimodal system, providing diverse information. Such optical systems with off-the-shelf optoelectronics can lead to affordable and improved biomedical tools. The aim of this thesis is to explore and demonstrate the applicability of CMOS SPAD line sensors with spectral photon counting and massively parallel timing capability in fluorescence spectroscopy, Raman spectroscopy and spectral reflectometry. For this, two SPAD line sensors are to be deployed, Ra-I and Ra-II, which were designed in the CMOS Sensors and Systems Group at the University of Edinburgh, led by Professor Robert Henderson.

The specific aims are the following:

- Characterise the Ra CMOS SPAD line sensors, with particular focus on timing characteristics.
- Build a fully functioning system for time-resolved fluorescence spectroscopy and Raman spectroscopy with Ra-I, including optoelectronics and software.
- Demonstrate the high photon processing capability of Ra-I in time-resolved fluorescence spectroscopy, and evaluate the efficiency of the system with respect to the acquisition speed of time-resolved and spectral fluorescence data with given spectral and temporal accuracy.
- Demonstrate the benefits of time-resolved capability in Raman spectroscopy.
- Build an electro-optical system for spectral-domain optical coherence tomography (OCT) with Ra-II, including hardware and software.
- Analyse and characterise SPAD detectors and the effect of detector parameters on OCT performance.
- Demonstrate the combination of time-resolved single-photon counting and low-coherence interferometry for advancing spectral detection of scattered light, and identify limitations of the technique at the current level of technology.

1.6 Thesis outline

The rest of the thesis is divided as follows.

Chapter 2: Introduction to Spectroscopic Techniques in Biomedicine

This chapter introduces three particular spectroscopic applications, namely fluorescence spectroscopy, Raman spectroscopy and OCT. This work studies how time-resolved SPAD arrays can be deployed in these three techniques to make improvements on the state-of-the-art. While the first two techniques are based on spectral detection of light, the third one, optical coherence tomography is slightly different in this regard. In OCT, different methods exist for recording interference signals from the sample of interest, based on time-domain or spectral detection. This thesis focuses on spectral-domain (SD) OCT, where a spectrometer setup can be deployed using a linear sensor array.

For each of the three techniques the working principles are discussed, crucial components and related requirements are identified, and applications using these techniques are introduced. The chapter also introduces the additional computational requirements and data processing to produce the desired information from the measurement data.

Chapter 3: Characterisation of CMOS SPAD Line Sensors

In this chapter, the characterisation methods of CMOS SPAD sensors are described. Two CMOS SPAD based line sensors are characterised: Ra-I and Ra-II, which were used in this work. The sensors are assessed regarding crucial parameters affecting their application, such as the noise level, timing of photons and detection uncertainty and means of temporal discrimination of optical signals. The methods for characterisation and respective results are discussed in details.

Chapter 4: Time-resolved Spectroscopy with Ra-I

This chapter focuses on advancing time-resolved fluorescence and Raman spectroscopy using the first generation of CMOS SPAD line sensor designed in our group (Ra-I) with massively parallel timing capabilities. The hardware, firmware and software surrounding the Ra-I sensor all contribute to high efficiency of time-stamping photons, high data-rate and ease of use for end users, these are thoroughly described. In time-resolved fluorescence spectroscopy, three example applications are demonstrated: measuring the kinetics of the spectral fluorescence signal of chlorophyll in an intact leaf, measuring the dynamic fluorescence response of a biosensor probe, and measuring the steady state fluorescence spectrum of excised mouse lung tissue. These examples are to illustrate the applicability of the system for fast acquisition of spectrally and temporally resolved fluorescence data. Additionally, initial results are shown for improved detection of Raman spectra using the time-gating feature of the Ra-I sensor.

Chapter 5: Spectral-Domain Optical Coherence Tomography with Ra-I and Ra-II

This chapter investigates the feasibility of OCT imaging with CMOS SPAD arrays. For the first time, spectral single-photon counting is demonstrated in an SD OCT setup, using pulsed illumination and SPAD line sensors. The effect of the sensor parameters of Ra-II on OCT is analysed through simulations and measurements. The benefits and drawbacks are identified and discussed in comparison with a commercially available CMOS camera.

Chapter 6: Time-resolved Spectral-Domain Optical Coherence Tomography

The main advantage of SPAD technology, massively parallel timing of photons is introduced as a novel detection concept when combined with low-coherence interferometry. The proposed technique and related requirements are described on a conceptual basis, with indications of potential applications. A proof of concept study is presented using the proposed combined detection technique to remove reflection artefacts in SD OCT.

Chapter 7: Summary and Conclusion

Finally, the results are summarized. Crucial characteristics of SPAD line arrays are discussed, focusing on properties where CMOS SPAD sensor designs need improved performance in the presented applications. We conclude by looking into future applications and several ideas which are to be followed up on.

Chapter 2

Introduction to Spectroscopic Techniques in Biomedicine

The detection and analysis of light signals is an indispensable tool in biomedicine. As mentioned previously, most of what we know in biology comes from optical instruments such as microscopes. Light can be absorbed, scattered, reflected or polarized to mention a few mechanisms. On the other hand, light needs to be generated or emitted either at the light source or from the sample itself (as is the case with fluorescence). This thesis focuses on three different time-resolved spectroscopic techniques analysing light emanating from tissue or biomedical samples. The first one, fluorescence spectroscopy is based on emission of light as a result of (pulsed) laser excitation (Lakowicz, 2006). In Raman spectroscopy, the characteristic scattering of the incident light on molecules of interest is considered (McCreery, 2000). The third technique, optical coherence tomography (OCT), is used to reveal micrometre resolution images of structures of scattering media such as biological tissues, utilizing the backscattered light from the sample (Drexler and Fujimoto, 2015). The common feature of all three techniques is that they require the measurement of spectral light intensities. Throughout the thesis, instrumentation and experiments are discussed where recording the temporal evolution of the spectral intensities at several hundreds of channels with pico and nanosecond resolution offers improvements on the techniques, overcoming associated challenges and offering novel approaches. To that end, the reader is introduced to these techniques from both steady-state and time-resolved perspectives.

2.1 Fluorescence spectroscopy

Fluorescence spectroscopy is one of the most ubiquitously applied spectroscopic techniques, used across biochemistry, biophysics and medical sciences. From the first reported observation of a vivid blue luminescence of quinine in 1845 by Frederik William Herschel when exposed to sunlight, fluorescence spectroscopy has expanded enormously to become a routinely used technique. Today, fluorescence spectroscopy is utilized in medical diagnostics, cellular and molecular imaging, detection of single molecules, measuring the size and shape of proteins and recording their conformational changes, to name only a few applications. Pulsed light sources and advanced sensors can be used to characterise the fast fluorescence decay after excitation (lifetimes of most fluorophores used in fluorescence microscopy are 0.5 to 5 ns). This leads to a rich set of results from single molecule fluorescence studies (Moerner and Fromm, 2003) and protein folding (Michalet *et al.*, 2006) to tissue auto-fluorescence in medicine (Marcu, 2012). In this section, the basic principles of fluorescence spectroscopy are introduced. Time-correlated single-photon counting (TCSPC) as an efficient technique for time-resolved spectroscopy is described, in consideration with the advancement that massively parallelised time-resolved single-photon avalanche diodes (SPADs) offer. Finally, the data processing challenges of time-resolved spectroscopy are discussed along with proposed solutions.

2.1.1 Introduction

Fluorescence

When fluorescent materials absorb light of a certain wavelength their atoms are excited and gain energy. Some of the absorbed energy is lost as heat, almost instantly. The remaining surplus energy is rapidly re-emitted as light of a longer wavelength (different colour) during relaxation of the excited atoms.

As discussed in chapter 1, electrons can be excited and promoted to higher energy orbitals by photon absorption. Fluorescence is a luminous phenomenon, which happens when an excited electron of a molecule returns to its ground state. During the ‘downward’ transition to ground state, a photon is emitted with an energy matching the energy difference of the two states. The term fluorescence is used when the excited electron occupies a singlet state, meaning that it has opposite spin state than its pair electron on the ground state. The average time spent in the excited state in this case is typically in the nanosecond range. There are also several other radiationless processes modifying the energy of the electrons, and various excited states, such as vibrational and triplet states. The electronic states and molecular processes causing transitions between them are best described on a Jablonski diagram (Fig. 2.1).

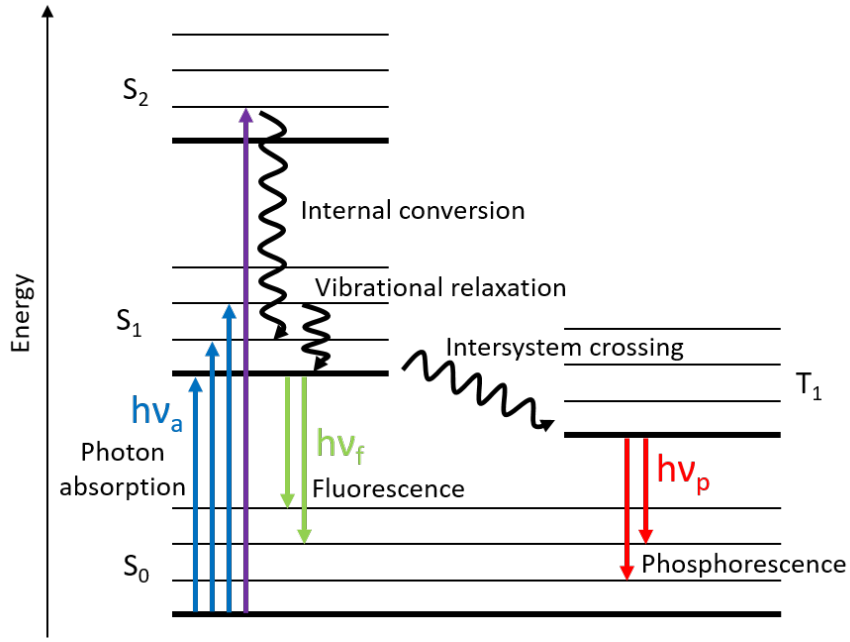

Figure 2.1: A typical Jablonski diagram depicting ground (thick horizontal lines) and higher vibrational states (thin horizontal lines) for singlet electronic states S_0, S_1 and S_2, and triplet electronic state T_1. Radiative transitions, involving photon absorption and emission, are shown with straight vertical lines. Non-radiative transitions, including internal conversion, vibrational relaxation and intersystem crossing, are shown with curly arrows.

When excited, the electron often ends up in a higher vibrational state. Some of the absorbed energy is turned to lengthening and shortening bonds, widening and squeezing bond angles, or lost in collisions and as heat. Due to the vibrational relaxation and other reactions, such as energy transfer and solvent-effects, fluorescence photon emission usually takes place from the lowest vibrational level of the excited electronic state. When returning to the ground electronic state (S_0 in Fig. 2.1), the electron usually returns to en excited vibrational level first, and then reaches the ground level after vibrational relaxation. Such losses between excitation end emission causes the emitted fluorescence photons having a longer wavelength as the excitation, known as Stokes shift. In most cases, especially when there are many vibrational levels and the excited electronic states are close so that their vibrational levels overlap, the rapid relaxation after excitation occurs over several electronic states (internal conversion). The result is that photon emission happens almost exclusively from the ground level of S_1 (Fig. 2.1) and consequently the emission spectrum is independent of the excitation, this is known as Kasha's rule (del Valle and Catalán, 2019).

During internal conversion, the excited electron may undergo spin conversion as well (especially in the case of heavy atoms). This process is called intersystem crossing,

and the system ends up in a triplet state (T_1 in Fig. 2.1). According to the Pauli exclusion principle, two electrons can only occupy the same orbital if they have opposite spin states. Consequently, the transition to the ground state from a triplet state is forbidden, and the electron is trapped in this intermediate state between the excited and ground states. The electron eventually drops back to the ground state with a rate several orders of magnitude lower than that of fluorescence. The average time spent in this state is much longer accordingly, typically meaning micro- and milliseconds, but up to several minutes in some cases. Photon emission from the triplet state is called phosphorescence. As the triplet state is lower in energy than the singlet state, phosphorescence emission has a longer wavelength.

Fluorescence emission has a characteristic spectrum depending on the chemical structure of the material under study and its solvent. Due to spectral broadness of fluorescence emission (and presence of many molecules fluorescing), it is often hard to disentangle which exact molecule is emitting the fluorescence, and therefore chemical fingerprinting is usually left to techniques such as Raman (McCreery, 2000, chapter 1 and 2) or infrared spectroscopy (Movasaghi *et al.*, 2008). However, when the sample is labelled specifically such as in smartprobes (Mills *et al.*, 2016), the specificity of detection rises fundamentally. As with other spectroscopies, the emission and absorption spectra are expressed in function of wavelengths (nm) or wavenumbers ($\frac{1}{\text{cm}}$).

Lifetime

Another characteristic feature of fluorescence is its lifetime. Relaxation of an excited electron and the emission of a fluorescence photon is a random process. However, a population of excited fluorophore species can be described quantitatively on a macroscopic level. If no continuous excitation is provided (i.e. using a short light pulse for excitation), the excited fluorophore population will eventually decay away. The speed of the depopulation process depends on the decay rate constant, which expresses the average number of transitions from the excited state to the ground state, projected to a single molecule and unit time (therefore having the unit of $\frac{1}{s}$). In the simplest case, an excited electron either releases a fluorescence photon (emissive or radiative decay), or loses the absorbed energy as heat (non-radiative decay). Both of these processes contribute to the depopulation of the excited state and the total decay rate (equation 2.1, where k_r, k_{nr} and k are the radiative, non-radiative and total decay rates, respectively). As a result, the excited population follows an exponential decrease in intensity with time. The exponential decay is characterized by the lifetime (τ), which is the reciprocal of the total decay rate and expresses the average time that a single electron spends in the excited state (equation 2.2) (Berezin and Achilefu, 2010). It should be pointed out, that when multiple fluorophores are present in a sample

under study, the excited population shows a multiexponential decay, and the measured lifetime is a weighted sum of the lifetimes of all contributing fluorophores. Also, the decay of the excited population may not be exponential at all in an inhomogeneous molecular environment and when the decay rates are not constant.

The rate of radiative and non-radiative relaxation processes also determine the efficiency of fluorescence emission. This is expressed by the quantum yield (Q), which is the ratio of the emissive decay rate and the total decay rate (equation 2.3). Equivalently, quantum yield can be defined as the ratio of the number of emitted photons and absorbed photons.

$$k = k_r + k_{nr} \tag{2.1}$$

$$\tau = \frac{1}{k_r + k_{nr}} \tag{2.2}$$

$$Q = \frac{k_r}{k_r + k_{nr}} \tag{2.3}$$

Measuring the lifetime has several benefits over steady-state measurements. The lifetime is an intrinsic property of fluorophores and as such, it does not depend on the intensity of fluorescence or the excitation, therefore it is not corrupted by low fluorescence signal levels or low fluorophore concentration. On the other hand, lifetime (especially the non-radiative decay rate) is sensitive to the internal structure of fluorophores and the environment of molecules such as temperature, pH and the presence of fluorescence quenchers. Changes of the fluorescence lifetime reveal molecular interactions, fluorophore-solvent reactions, conformational changes and Förster resonance energy transfer (FRET) (Jares-Erijman and Jovin, 2003). FRET takes place when two molecules are close in space, and the fluorescence emission spectrum of one (donor) overlaps with the absorption spectrum of the other (acceptor). Energy transferred from the donor to acceptor and the associated change in lifetime provide means to measure microscopic distances and the spatial distribution of donor and acceptor molecules. Lifetime measurements are also favourable to distinguish between labels with overlapping emission spectra in biomedical imaging.

2.1.2 Time-resolved fluorescence measurements

As suggested by the importance of fluorescence lifetime, time-resolved fluorescence measurements have major benefits. On the other hand, time-resolved measurement techniques are inherently more complex to implement. Fluorescent lifetimes are usually in the nanosecond or sub-nanosecond range. Measuring such short time intervals demands sophisticated and expensive instruments, advanced light sources and high-speed detectors in particular. Time-resolved detection can be classified to time-domain and frequency-domain measurements. The latter one is briefly mentioned in the previous chapter, but not discussed in details as part of this thesis. A detailed description of frequency-domain (FD) fluorometry is given by Lakowicz (2006, chapter 5).

Time-domain (TD) techniques aim to derive the fluorescence lifetime from the recorded temporal change of fluorescent emission intensity. Since this is proportional to the excited population, the lifetime can be approximated from the slope of the measured intensity decay (Fig. 2.2). For time-domain measurements a pulsed light source is used to excite the sample. The ideal excitation pulse is a Dirac delta function (a pulse with zero width but finite integral). In practice, the pulse width is reasonably shorter than the expected lifetime. Typical light sources with ps and fs pulse widths and MHz and GHz repetition rates are mode-locked lasers, supercontinuum lasers, solid-state pulsed laser diodes and light-emitting diodes (LEDs). Following excitation, the emitted fluorescence intensity is recorded against time. The ideal detector turns all incident photons to electric signals instantly, and it is ready for photon detection any time. In reality, detectors such as photomultiplier tubes (PMTs), microchannel plates (MCPs) and photo diodes all have finite detection efficiencies, timing uncertainties and deadtimes. The fluorescence lifetime is estimated from the measured decay curve. Usually, the recorded curve is distorted by the finite excitation pulse width and timing uncertainties introduced by optics and the detector. Robust algorithms take the detrimental effect of these instrumental factors into consideration. This requires the measurement of the instrument response function (IRF), which estimates the response of the whole measurement apparatus to a hypothetical sample with zero lifetime. The processing steps for retrieving the lifetime is described in detail in the following sections.

Several different methods have been proposed for measuring the temporal profile of fluorescence emission intensity. The most intuitive method is the direct measurement of the decay in time with a fast detector (Marcu *et al.*, 2014, chapter 4). In pulse sampling measurements, the time-dependent output of the detector is sampled with a fast digitizer or a high-end oscilloscope. The achievable timing resolution in these measurements is limited by the sampling speed. Furthermore, inherently low optical signals impose difficulties in direct measurements. Therefore, the direct measurements usually require low repetition rate, but high peak power lasers such as CO_2

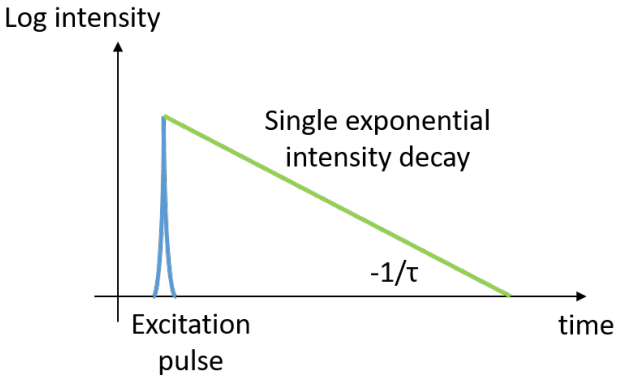

Figure 2.2: Time-domain measurement of fluorescence lifetime. A light pulse excites the fluorophore which emits photons with decreasing intensity according to the decay of the excited molecule population. Provided a single exponential decay, the lifetime (τ) is given by the slope $(-\frac{1}{\tau})$.

lasers (Marcu, 2012) which, in turn, generate enough signal for direct sampling. However, an excessive amount of energy of the incident light can cause photochemical destruction of the molecules in the sample. The consequence of this effect, called photobleaching, is a reduced fluorescence emission from the sample.

Gated acquisition of the emission signal can also be applied to record the waveform through several emission cycles (Marcu *et al.*, 2014, chapter 8; Becker, 2005, chapter 2). In this technique, a time-gate narrower than the emission decay is applied upon detection. The gate is scanned over the decay, and a different part of the decay is measured at each emission cycle. Gating can be achieved electronically with image intensifiers, and gated photon counters. In the former case, gating is achieved by modulating the voltage applied to the intensifier (see section 1.2.2). Depending on its direction, the voltage either accelerates photoelectrons towards the phosphor screen, or prevents their propagation further in the device. In the case of gated photon counters, a single-photon sensitive detector is used with a counter, where counting of the photon pulses is only enabled for certain time periods. The counter enabling electronic pulses (gate pulses) are often triggered synchronously with a pulsed light source, with delay generators controlling the gate pulse width and position in time (Becker, 2005, chapter 2). Optical gating is performed with the use of nonlinear photonic crystals. These materials exhibit a nonlinear response to incident light, that allows the use of an optical gating signal to change the refringence of the material (optical shuttering), or shifting the wavelength of an incident light beam when overlapping spatially and temporally with the gating beam (upconversion) (Xu and Knutson, 2008). These methods provide high gating resolution but infer complex and expensive instrumentation. Gated measurements are particularly useful when the detector is limited in speed

or bandwidth for direct time-resolved measurements. The counting efficiency, on the other hand, is lower than that of a direct measurement. The counting efficiency is increased when multiple gates are used at different timepoints, at the expense of complexity and cost. The main limitation of this method is the number of applicable gates and the width of each gate compromising the temporal resolution. In both cases, the overall uncertainty of detection is increased with the jitter of the gate.

Detectors operating in single-photon counting (SPC) are used with multichannel scalers in other systems (Becker, 2005, chapter 2). These devices perform a fast sweep of memory locations synchronously to a reference signal (excitation), and sort the detected photons to the memory locations accordingly. The achievable time-resolution of these solutions is around 1 ns, which is still not sufficient for fluorescence measurements of short lifetimes. TCSPC offers the highest flexibility and lowest noise (Alfano and Ockman, 1968). This thesis mainly focuses on TCSPC for time-resolved measurement of fluorescence decays.

2.1.3 Time-correlated single-photon counting

TCSPC is a technique widely applied in photonics, solid-state physics and nuclear physics (Becker, 2005, chapter 5) for recording the timing of discrete events with respect to periodic excitation. For example, in the case of fluorescence the arrival time of single fluorescence photons is recorded with respect to the excitation pulse that generated the fluorescence emission. When measuring picosecond or nanosecond events, one needs to resort to high resolution timing circuits with core time interval measured ideally much lower than the event sampled (i.e. 50 ps time resolutions to measure 2 ns fluorescence decay). The working principle is best described by the following. The timing circuits are started when an excitation pulse is produced. The emission signal in response is not measured directly, instead the signal intensity is maintained such that a maximum of one emission photon is detected. For strong fluorescence emission, this may require attenuation of the signal. Upon detection of this photon, the timing circuit is stopped, and the measured time-of-arrival is noted (Fig. 2.3a,b). The timing circuit is reset and the timing measurement is repeated with each excitation pulse. Subsequent measurements of arrival times are used to create a histogram where the number of occurrences of each possible arrival time (based on the timing resolution) is registered. Under certain conditions regarding the emission count rate, the number of different arrival times (i.e the histogram) follows the emission intensity profile (Fig. 2.3c).

In practice, the reference event starting the time measurement (START) is provided by the sync signal of the pulsed light source (or given as the output of a separate detector sensing the excitation pulses). In TCSPC, the count rates are relatively low.

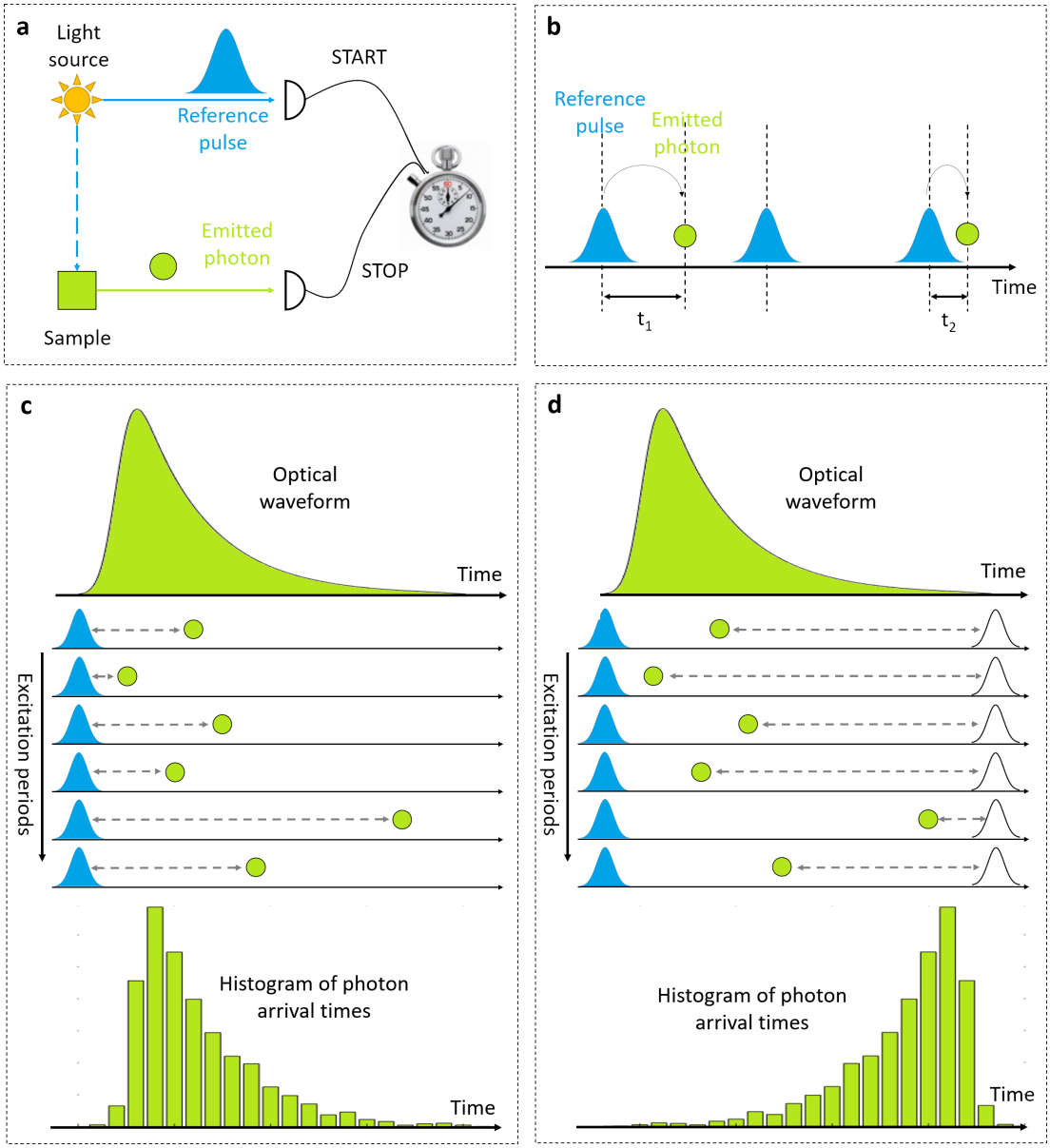

Figure 2.3: In time-correlated single-photon counting the sample is excited with a short light pulse, and photons of the fluorescence emission are detected (a). The arrival time of single emission photons is recorded with respect to a reference event, provided by the excitation pulses (b). Subsequent measurements of photon arrivals are collected in a histogram. Provided a low count rate of emission photons, the generated histogram follows the true shape of the optical waveform of fluorescence emission (c). In reversed START-STOP mode the timing circuits are started by the emission photons and stopped by a delayed excitation pulse, so there is no need for resetting the timers when no photon is detected in the laser cycle (d).

Consequently, the timing circuits need frequent resetting with no detected photons stopping them in most excitation cycles. This was particularly problematic with time-to-analogue converter (TAC) with long reset times, decreasing the duty cycle of the timing circuits and requiring low repetition rate of the excitation. As a solution, recent TCSPC instruments all apply a 'STOP-START' or reversed 'START-STOP' operation. In this operation mode, photon detection embodies the START event (instead of stopping the timing). The timing circuits are stopped by a delayed (or a subsequent) sync pulse. As a consequence, the recorded histograms are flipped along the time axis (Fig. 2.3d). Delay of the sync signal shifts the entire histogram along the time axis. While, in theory, this does not change the shape of the measured signal, there are a few points worth mentioning. Timing in time-to-digital converters (TDCs) is often carried out by a periodically changing electrical signal. The jitter of each repeated timing stage adds up and can be significant if the timing circuit is running for a long time. Therefore, the delay in the sync signal should be carefully aligned so that the timing circuit runs for short times. The spacing of the electrical sync pulses (and the optical emission pulses) is set by the repetition rate of the laser, and therefore it is assumed that the temporal distance between pulses is constant. Still, non-corresponding excitation and emission pulses might be less correlated (i.e. when the timing is stopped by a different reference event than the one initiated the emission that started the timing). The result is a higher uncertainty of timing when the signal path of the optical (emission event) or the electrical (reference event) signals is significantly longer.

TCSPC requires short excitation pulses. The first light sources were dye lasers, argon ion and Ti:sapphire lasers with ps and fs pulses. The appearance of simple and cost-effective sources, solid state laser diodes and LEDs, has had a major contribution to the accessibility and spread of TCSPC systems. These sources are available at a wide range of wavelengths in the visible and near infrared (NIR) range, they can be pulsed at variable repetition rates (GHz for laser diodes and 10s of MHz for LEDs), and provide short pulses (ps for laser diodes and ns for LEDs). The detector needs to have single-photon sensitivity. While the speed of conversion from photons to electric signals is often an important parameter, long conversion times (e.g. with PMTs) are not limiting, since they only add a delay to the timing of photons. The uncertainty of the detection time (transit time spread or jitter) is crucial, however. Timing uncertainties are introduced by various effects, depending on the sensor technology. As an example, with SPADs, the detection time depends on statistical fluctuations of the location where charge carriers are generated within the device and their diffusion to the depleted region, but also on the spread of the avalanche current. In the case of PMTs, the sources of jitter include the location and size of the illuminated region of the photocathode, the location of electron emission on the electrodes, the initial velocity of the emitted electrons and fluctuations in the electron trajectories. The traditional devices

of choice have been PMTs and MCPs with high gain and low timing jitter. SPADs offer similar jitters (tens of ps) as MCPs at a lower cost and compact construction. PMTs and MCPs are analogue devices. Sensing the onset of the analogue detector output increases the complexity of the TCSPC electronics chain and the timing uncertainty of detection. Photon detection on SPAD sensors is a digital event, which can be directly processed with integrated timing circuits.

When the photon count rate is low, several detector channels could be used with a single timing circuit for multidimensional TCSPC (Becker, 2005, chapter 3). This allows the measurement of several decays at different locations at the same time, or time-resolved spectral fluorescence measurements. While streak cameras are costly and complex detectors, they allow the detection of spectral fluorescence decays at several 10s of channels with picosecond time resolution (10 ps in practice). The additional benefit is that they are not limited to low light levels associated with TCSPC operation. Multidimensional TCSPC is possible with PMT arrays and position sensitive PMTs with compact TCSPC electronics on Peripheral Component Interconnect (PCI) cards. These solutions are limited to 16 channels. Complementary metal-oxide-semiconductor (CMOS) technology allows system on chip (SOC) solutions with several hundred individual TCSPC channels with SPAD sensors integrated on a single chip.

Counting loss, pile-up and saturation

Light levels in TCSPC are required to be low due to the deadtime of TCSPC electronics. The timing circuits are unable to detect further photons for at least several 10s of nanoseconds after detection of a photon. If multiple photons arrive in a single excitation cycle, only the first one is detected (in most TCSPC systems). This leads to lost photons (counting loss), and over several cycles to an over-representation of early photons and distortions of the recorded histograms consequently. The distortion of the measured intensity decay is minor when the detected count rate is up to 10 % of the excitation repetition rate (i.e. 1 emission photon in every 10 excitation cycles). Even if the count rate is limited so that there is a high probability of receiving only one photon per emission cycle (maintaining the $<$10 % rule), no photons should arrive during the deadtime of the device either. A practical limit is expressed by the 'useful count rate', which is 50 % of the reciprocal of the deadtime. Limitations regarding the excitation pulse rate should be considered as well. The pulses should not be closer than ~4 times the fluorescence lifetime in order to avoid re-excitation of the decay. Correction methods have been proposed to reduce the effects of deadtime and pile-up, based on processing of the measured data (Becker, 2005, section 7.9) using statistical assumptions. The most efficient method to prevent pile-up and increase the counting

capability is to use a multiplexed timing system with several channels (Arlt *et al.*, 2013).

When several timing channels are used, the length of the exposure time becomes important as well. The exposure time does not affect the recorded temporal profiles, however it may distort the information measured across the channels. Let us consider a spectral TCSPC system, where fluorescent decays at different wavelengths of the emission spectrum are measured at different channels. To build decay histograms, the sample is repeatedly excited according to the TCSPC operation principles. In practice, the detector measures the emission signal over a certain exposure time covering many excitation cycles. At the end of the exposure time, the measured photon arrival time is provided by the timing circuits at each channel (if at least one photon is detected on the channel). Let us further assume that the emission signal is attenuated such that only a single photon is detected for many excitation pulses ($>$10), to avoid early photon pile-up. If the exposure time is long enough, no matter how low probability the detection of a single-photon has, one photon will be detected sooner or later on all channels. In this case, the number of photons in the histograms of each channel will be the same, independently of the different spectral intensities falling to each spectral channel. In other words, the steady state intensity calculated from TCSPC data has a maximum value set by the number of exposure cycles used for building the TCSPC histograms. Consequently, for exposure times too long, the integral of the of the spectral channels along time would look like as a clipped steady state spectral intensity. To avoid such saturation of a channel, the exposure time needs to be set so that the average number of photons detected during exposure time ($< N_{photon} >$) is proportional to the intensity of the emission signal on that channel. That is, the number of excitation pulses (L) per exposure time needs to be lower than the reciprocal of the highest count rate (R), expressed in number of photons per excitation pulses (see Fig. 2.4). Alternative timing solutions were reported to record multiple photon arrivals in a single laser excitation cycle (Dutton *et al.*, 2014; Erdogan *et al.*, 2017; Al Abbas *et al.*, 2018).

$$< N_{photon} >= \begin{cases} RL : L < \frac{1}{R} \\ 1 \quad : L \geq \frac{1}{R} \end{cases}$$

2.1.4 Processing and analysis

Weighted nonlinear least squares

The lifetime (or lifetimes) needs to be derived from the measured time profile of the fluorescence decay. This is most often done by weighted nonlinear least squares (LS) fitting. Generally, the task of the fitting is to find the parameters ($\boldsymbol{p} = [p_1, p_2, ..., p_{N_p}]$) of a parametric mathematical model ($f(x, \boldsymbol{p})$, e.g. an exponential decay with $p = \tau$

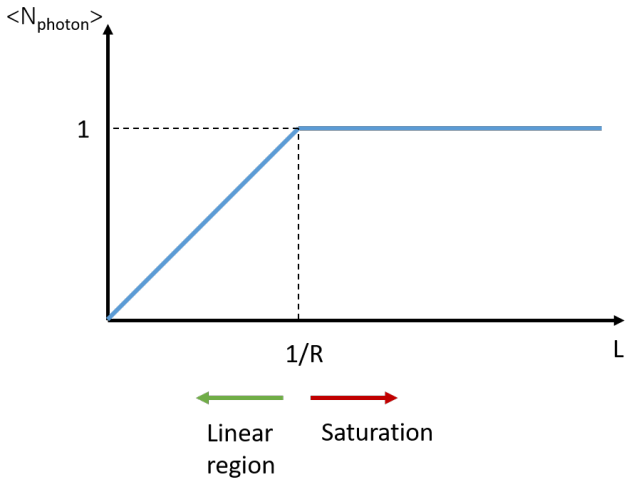

Figure 2.4: The number of photons detected in an exposure cycle is limited to 1 in time-correlated single-photon counting (TCSPC). Theoretically, the intensity (number of counts) in a TCSPC experiment is linear with the true intensity, if the exposure time is short enough. $< N_{photon} >$: expected number of photons per exposure cycle, R: expected number of photons per laser cycle (e.g. 0.1), L: number of laser cycles per exposure cycle.

lifetime) for which the model function is 'closest' to the data points of the measurement ((x_1, y_1),(x_2, y_2),...,(x_{N_x}, y_{N_x}), $N_x \geq N_p$, e.g. measured intensity y_i of the decay at timepoints x_i). The 'closeness' of the model with a certain set of parameters is quantitatively measured by the weighted summed square of residuals (r_i), called the χ^2 value (see equation 2.4).

$$\chi^2 = \sum_{i=1}^{N_x} \frac{1}{\sigma_i^2} {r_i}^2 = \sum_{i=1}^{N_x} \frac{1}{\sigma_i^2} (y_i - f(x_i, \boldsymbol{p}))^2 \tag{2.4}$$

$$\chi_R^2 = \frac{\chi^2}{(N_x - N_p)} \tag{2.5}$$

Weighting of squared differences between the measured data and the prediction is required, since the observed decay values (in function of time) are not equally reliable. This is because the light levels (number of photons) have Poisson statistic, and the variance of each observation is equal to the measured intensity. Weighting with the measured light intensity (i.e. the variance of the observation) provides that each intensity measurement has equal contribution to the goodness of the fit. Starting with an initial set of parameters, minimum χ^2 is searched for iteratively using optimization algorithms, such as the Gauss–Newton method, the Levenberg–Marquardt algorithm and the Nelder–Mead method. Since the χ^2 value depends on the number of meas-

urement points (N_x), often the reduced χ^2_R value is applied (equation 2.5). For a perfect fit, χ^2_R is equal to 1. Different techniques have been proposed to evaluate the uncertainty of the derived parameters and associated confidence intervals (Becker, 2005, section 4.9 and 4.10). Nonetheless, a visual confirmation of the residuals often reveals if the residuals are not randomly distributed, but show systematic errors. In these cases, a more complex model may be required, even though increasing the complexity of fitting. Resolving multiexponential decays is generally challenging as the amplitude and lifetime parameters of the exponential components are highly correlated. In practice, there is a limit with respect to the measurement noise above which components with similar lifetimes cannot be distinguished (van den Bos and Swarte, 1993). For this reason, measurements of multiexponential decays require high signal-to-noise ratio (SNR).

Other analysis methods

Least squares regression assumes that the observations are normally distributed. With low photon counts of the decay, the measurement values are more accurately described with Poisson statistic. In this case, more accurate estimation of the lifetime is possible with maximum likelihood estimation (MLE) (Tellinghuisen and Wilkerson Jr., 1993). Here, the lifetime (or lifetimes) is an unknown parameter of a statistical model, in which each observation of the fluorescence intensity at different timepoints is affected by Poisson noise. MLE estimates the lifetime parameter(s) of this model, for which the observed data is most probable. Several other methods exist for the retrieval of the characteristic timing information of the fluorescence decay, which are less frequently applied. The centre of mass technique, for example, is less demanding computationally and therefore faster than fitting routines. However, it is also less accurate (Li *et al.*, 2010). This method can be sufficient when simply a differentiation between known fluorophores is required instead of accurate resolving of complex decays. Lifetime estimation can be improved through corrections with the IRF and the use of lookup tables (Krstajić *et al.*, 2015; Poland *et al.*, 2016). The applied method also depends on the mode of data acquisition. As an example, the decay might be undersampled when time-gated detection is used with low number of gates. With the *a priori* information of the decay shape, a few number of time-gates are sufficient for lifetime determination using the rapid lifetime determination (RLD) method, e.g. two gates for a single exponential decay, or four gates for a double exponential. RLD is inferior to fitting algorithms with respect to accuracy (30-40% worse precision compared to weighted least squares calculation with 512 measurement points (Ballew and Demas, 1989)), and it is unsuitable for resolving complex decays. However, for simple single and double exponential decays its simplicity allows substantially faster lifetime estimation (Sharman *et al.*, 1999).

Correction of the instrument response function

The contribution of time spread in the system (instrument response function or IRF) needs to be taken into account when processing the measured data. The IRF can be thought of as the impulse response function of the measurement system (response to a Dirac delta function, i.e. infinitely short pulse). The total IRF of a TCSPC system ($IRF(t)$ in equation 2.6) is the convolution of the time response of the light source ($u(t)$) and the timing accuracy of the electro-optics ($IRF_o(t)$) and the detector ($IRF_d(t)$). Timing resolution of the system is usually described as the full width at half maximum (FWHM) of the IRF. To measure the total IRF, scattering solutions are used as a sample illuminated at the expected wavelength of the fluorescence emission under study. The fluorescence signals ($x(t)$ in equation 2.7) are convolved with the IRF ('physically') when measured with the system ($y(t)$), therefore the applied decay model is convolved with the measured IRF (numerically) before fitted to the measured data during its analysis. Equation 2.8 describes how one would define the model ($f(t,\tau)$) assuming a single exponential decay, from which the lifetime value (τ) is to be estimated by fitting the model to the measured data ($y(t)$) (e.g. as described in equation 2.4). Owing to this reconvolution step, it is possible to measure decays with lifetimes shorter than the IRF (chapter 2 of (Becker, 2005)).

$$IRF(t) = u(t) * IRF_o(t) * IRF_d(t) \tag{2.6}$$

$$y(t) = x(t) * IRF(t) \tag{2.7}$$

$$f(t,\tau) = Ie^{\frac{-t}{\tau}} * IRF(t) \tag{2.8}$$

2.2 Raman spectroscopy

2.2.1 Introduction

Raman spectroscopy is a technique to study vibrational and rotational modes of molecules in a sample (a general introduction is given for example by Atkins *et al.* (2013, chapter 19)). Since these properties are very specific to the molecules in question and their environment, Raman spectroscopy has become a versatile tool for chemical analysis and identification of elements. The technique is based on the detection of light from a high intensity source inelastically scattered on the molecules of study. When the sample is illuminated, the majority of scattered light undergoes Rayleigh scattering. This means an elastic collision of photons with the sample molecules, where the momentum and the kinetic energy of the system is preserved. The Rayleigh

scattered photons do not lose kinetic energy, and their wavelength stays the same as the excitation wavelength. One out of 10^6–10^8 photons suffers inelastic collision, that is Raman scatter, and its kinetic energy changes. In a simple explanation, this is because the molecules can vibrate in different ways, depending on their orientation, bonds and the mass of their atoms. Each mode of vibration is associated with a specific frequency and energy, accordingly. The energy of the incident photons changes when they promote the molecule to a vibrational mode of higher or lower energy. The wavelength of these photons changes respectively. The phenomenon can be described using virtual energy levels (Fig. 2.5). In response to the incident light, the electrons do not move to a higher electronic state, but a virtual one. In the case of Rayleigh scattering, they return to the original vibrational level and the energy of the scattered light is not altered. In some cases the electrons return to a higher vibrational state than the one they originated from. Part of the energy of the scattered photons is taken away and turned into molecular rotations and vibrations. The case is called a Stokes shift, and the wavelength of the scattered light shifts according to the energy difference of the vibrational states. In other cases, when the electrons are on a higher vibrational level, they may return to a lower vibrational level upon interaction with a photon, increasing the energy of the scattered photon. The scattered light has a lower wavelength in this case than the excitation, and the case is called an anti-Stokes shift.

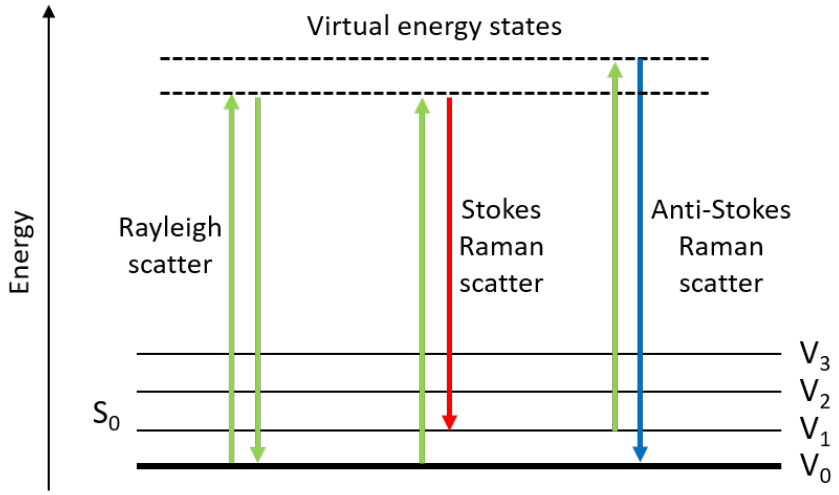

Figure 2.5: Energy-level diagram of Rayleigh scattering, Stokes and anti-Stokes Raman scattering. S_0 is the ground electric state where the molecule can exist at different vibrational states ($V_0, V_1, ...$)

The intensity of scattered light is usually plotted against the shift relative to the excitation. The shift is usually expressed in wavenumbers with a dimension of cm^{-1}:

$$\Delta\widetilde{\nu} = \left(\frac{1}{\lambda_{inc}} - \frac{1}{\lambda_{scatter}}\right) \tag{2.9}$$

where λ_{inc} and $\lambda_{scatter}$ are the wavelengths of the incident light and scattered light, and $\Delta\widetilde{\nu}$ is the shift expressed in wavenumbers. Since both Stokes and anti-Stokes scatter appear at the characteristic vibrational energy levels, intensity peak positions are symmetrical to the excitation in the spectral intensity plots (Fig. 2.6). However, anti-Stokes scattered light has much lower intensity, as the majority of molecules are at the lowest vibrational level when excited. For this reason, most Raman techniques study the Stokes scatter only. Coherent anti-Stokes Raman spectroscopy (CARS) (Evans and Xie, 2008) exploits anti-Stokes Raman scattering, but it is not widely applied in medicine for now. Raman spectroscopy is favourable as it is purely based on interaction with light at powers that are non-destructive, it is non-contact and requires no sample preparation. However, the technique has its own drawbacks as well. For example, high SNR spectra often require acquisition lengths of many seconds, as the photon budget in Raman is inherently lower than in fluorescence. Research efforts are therefore focusing on accelerating Raman spectroscopy acquisition.

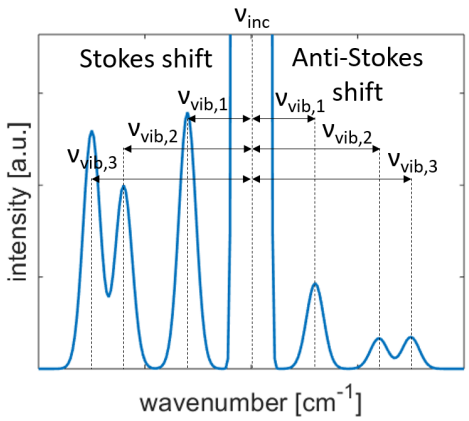

Figure 2.6: The intensity of scattered light in function of its wavenumber. The high peak in the middle (not fitting into the plot) belongs to Rayleigh scattered light, with ν_{inc} wavenumber. The Stokes and anti-Stokes Raman peaks appear symmetrically at characteristic wavenumbers belonging to the vibrational levels ($\nu_{vib,1}, \nu_{vib,2}, ...$)

2.2.2 Light sources and detectors

The main challenge in Raman spectroscopy is the detection of the weak Raman signal. In spontaneous Raman scattering the amount of Raman scattered photons is directly proportional to the excitation power. To this end, an excitation source with high power is needed. Unfortunately, this increases the amount of Rayleigh scattered light as well. Intense Rayleigh scattered light can easily damage the detector, but even at a tolerable level it does not contain any information. To remove Rayleigh scattered light, optical filters can be used, such as notch filters and bandpass filters. The linewidth of the light source is also essential. The Raman scattered light is shifted from the excitation

wavelength according to the vibrational and rotational energy levels. It follows from this working principle, that a large linewidth of excitation results in wide Raman peaks as well. When the small amount of Raman scattered photons are extended over a wider range of wavelengths, the Raman peaks spread and decrease and may fade into the noisy background. For the same reason, a high spectral resolution of the optical setup is also required. High power, monochromatic lasers are typical light sources of Raman instruments. The wavelength of the laser also affects the measurements in several ways. The most important effect to consider from this respect is the wavelength dependence of the efficiency of Raman scattering, which is inversely proportional to the fourth power of the excitation wavelength.

Systems applying lasers of lower wavelengths are therefore more efficient. However, in these cases, the energy of photons is comparable to the energy difference of electronic sates, and the induced fluorescence often overwhelms the weak Raman signal. Low intensity of Raman scattering requires high sensitivity, low noise detectors as well. Traditional detectors have been PMTs in Raman measurements. Measuring the spectrum of Raman scattering with single channel PMTs is time-consuming, therefore most systems apply spectral detection using charge-coupled device (CCD) and CMOS detectors. SPADs have been introduced to Raman recently, with the capability of time-resolved detection. Focusing the time of detection to the Raman peak only, and not recording the background noise outside the narrow time window increases the SNR of the measurements, which leads to easier detection of weak Raman peaks, and speeding up Raman measurements.

2.2.3 Advanced Raman techniques

Numerous variants have been proposed to overcome the limitations of Raman spectroscopy. Many of these techniques are focusing on increasing the amount of Raman scattering. In stimulated Raman spectroscopy (SRS) (Kukura *et al.*, 2007) an increased incident electric field is provided by pulsed illumination. The Raman intensity is increased through a non-linear process, but the increase is limited to certain vibrational modes only. Resonance Raman spectroscopy (RRS) (Efremov *et al.*, 2008) aims to enhance the Raman scattering through resonance effects arising when the excitation wavelength coincides with the electronic transitions of the studied molecules. This approach is limited to certain materials, and usually requires wavelength tunable lasers. The induced fluorescence further limits the applicability. Surface-enhanced Raman spectroscopy (SERS) (Stiles *et al.*, 2008) and tip-enhanced Raman spectroscopy (TERS) (Schmid *et al.*, 2013) also build on enhancement of the Raman signal through non-linear effects, although the enhancement is often specific to certain vibrational transitions. The resulted spectra often needs careful analysis and hence

the industrial use of these techniques is limited. In CARS (Tolles *et al.*, 1977) the anti-Stokes scattering is amplified through coherence. Other techniques are focusing on the reduction of fluorescence emission which hinders the visibility of Raman scattered light (Wei *et al.*, 2015). Shifted-excitation Raman difference spectroscopy (SERDS) (Shreve *et al.*, 2005) builds on the fact that in most cases the fluorescence emission is independent of the excitation wavelength over a certain range. In contrast, the Raman shift is always relative to the excitation. Two spectra are recorded with slightly different excitation wavelengths. Then, the purified Raman spectrum is reconstructed from the difference spectrum in which the fluorescence contribution is minimised. Time-gated Raman spectroscopy builds on separating the Raman signal and the fluorescence in time. Raman scatter is virtually instantaneous with the incident light, while fluorescence excitation is followed by vibrational relaxation first, then the emission has a decay with ns timeconstant. To maintain sharp Raman peaks in time and good separation of signals, a pulsed laser with short pulses is required. The detector is a time-resolved detector with high timing resolution and low jitter.

2.3 Optical coherence tomography

2.3.1 Introduction

Optical coherence tomography (OCT) is an optical imaging technique providing high-resolution, sub-surface, cross-sectional images of scattering media, such as biological tissues. Compared with other techniques, the resolution and penetration depth of OCT fit between confocal microscopy and ultrasound imaging. Confocal microscopy provides excellent submicron resolution, but a limited depth penetration of a few hundred microns. Ultrasound offers tens of centimetres of penetration, but the achievable resolution is just below a millimetre (Szabo, 2014). The resolution of OCT images is in the micrometre range, with a penetration depth of a few millimetres. Structures and surfaces along the depth (axial direction) are uncovered by measuring the magnitude and echo time delay of light of a source backreflected and backscattered from the sample. In OCT terms, this is called an A-scan. Two and three-dimensional images (B-scans) are generated by scanning the light beam in a transversal direction (Fig. 2.7). OCT allows live imaging *in vivo in situ*. In contrast to histopathology, which is time consuming and requires the removal and processing of tissue, OCT offers non-contact, noninvasive (or minimally invasive) optical biopsy. The features in an OCT image show up due to different optical scattering properties of structures in the sample, and therefore the samples do not require contrast agents, staining or other type of preparation. Internal body and intravascular imaging is allowed by integration with catheters, endoscopes and needles (Yaqoob *et al.*, 2006; Bezerra *et al.*, 2009). OCT

is used across a wide range of medical fields including ophthalmology, dermatology, cardiology and *in vivo in situ* optical biopsy of internal organ systems (Fujimoto *et al.*, 2000).

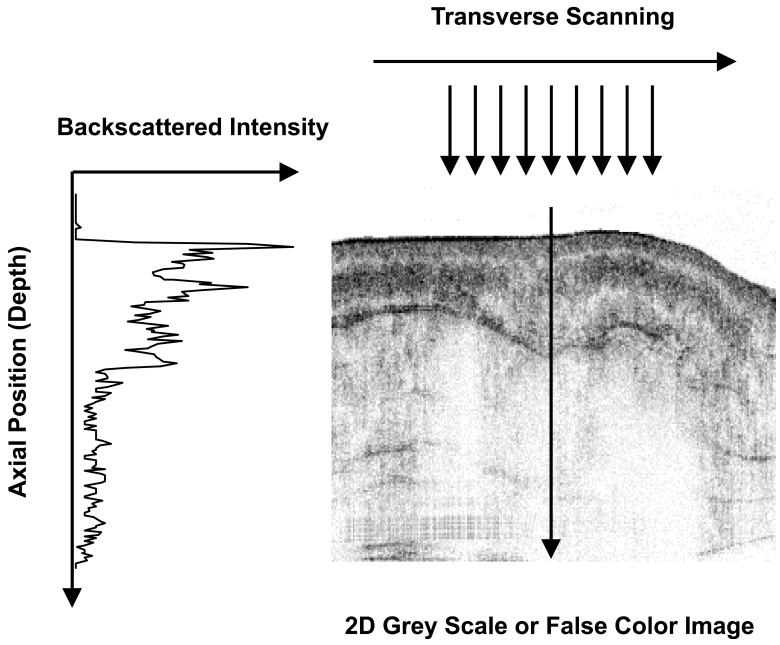

Figure 2.7: Image formation in optical coherence tomography. Backscattered intensity versus depth is measured on an A-scan (left), based on the echo time delay of light. Subsequent measurements at different transverse scanning positions reveal the structures along a cross-sectional plane. Figure from Fujimoto *et al.* (2000).

Time-domain optical coherence tomography

OCT technique is similar to ultrasound, which measures the echo time delay of sound waves reflected from structures of the studied medium. However, in OCT the information is carried by optical signals. For sufficient penetration to the tissue, the incident light is in the NIR region. Direct measurement of the echo time (time-of-flight (TOF) measurement) for micrometre resolution images would require time-resolved light detection with femtosecond resolution. Instead, low-coherence interferometry is applied. For this, broadband light is used for illumination. The light is split in a Michelson-interferometer, part of it is directed into the sample of interest, using one arm of the interferometer (sample arm, see Fig. 2.8a,c,d). The other part is backreflected from a mirror in the other arm of the interferometer (reference arm). Light backreflected and backscattered from surfaces and structures of the sample is superimposed in the interferometer with light of the reference arm. Due to the short temporal coherence of light, interference only takes effect if the path difference between the sample and reference arms is within the coherence length. In time-domain optical coherence tomography (TD OCT, Fig. 2.8a) the interference intensity is measured on a single point detector

while the reference arm mirror is scanned along the depth range (Huang *et al.*, 1991). The amount of interference detected at a certain reference mirror position depends on the backscattered intensity from a region of the sample, that has a depth matching the reference arm path length. The envelope of the measured signal, as the mirror is scanned, reveals the reflectivity profile along the depth (Fig. 2.8b).

Fourier-domain optical coherence tomography

In Fourier-domain optical coherence tomography (FD OCT) (de Boer *et al.*, 2017) the backreflected or backscattered light is collected from all depth levels of the sample at the same time. Instead of measuring the interference intensity while the reference mirror is scanned, the spectrum of the interference signal is measured with a fixed reference mirror. The depth profile is revealed from the spectrum of the interferogram, relying on the Fourier-transform relation of a signal's autocorrelation function and its spectral power density. The spectral interference can be measured by spectral detection, using a broadband light source and the interference signal dispersed onto a detector array (spectral-domain optical coherence tomography (SD OCT), Fig. 2.8c). In this case, the interference spectrum is retrieved from the detector array in a single measurement (Fercher *et al.*, 1995). Alternatively, single channel detection is used with a narrow linewidth of light swept across the spectrum (Chinn *et al.*, 1997) (swept-source optical coherence tomography (SS OCT), Fig. 2.8d). With a swept source instrument, the spectrum is measured as the detected signal intensity in function of time.

In both cases, each structure along the sample causes a sinusoidal intensity modulation (fringes) of the interference spectrum. The periodicity of the modulation is inversely related to the path difference between the respective structure in the sample and the reference mirror. The sum of these fringes with different modulation periodicities is what is seen on the measured interference spectrum (Fig. 2.8e top). Fourier analysis is applied to untangle components of the modulation with different periodicities, and to resolve reflections/structures at different depths, i.e. the depth profile (Fig. 2.8e bottom). The main advantage of FD OCT systems over time-domain optical coherence tomography (TD OCT) is a higher SNR, which scales with the number of spectral channels (de Boer *et al.*, 2003; Leitgeb *et al.*, 2003; Choma *et al.*, 2003). In FD OCT, no mechanical scanning of the reference arm is required. Instead, the axial scan rate is either set by the line rate of the detector array (SD OCT) or the sweep rate of the source (SS OCT). FD OCT provides high acquisition speed, enabling video-rate volumetric imaging and reduced motion artefacts.

It is not straightforward to recognise how TD OCT and FD OCT provide the same depth information. A detailed mathematical proof is given, for example, by Drexler

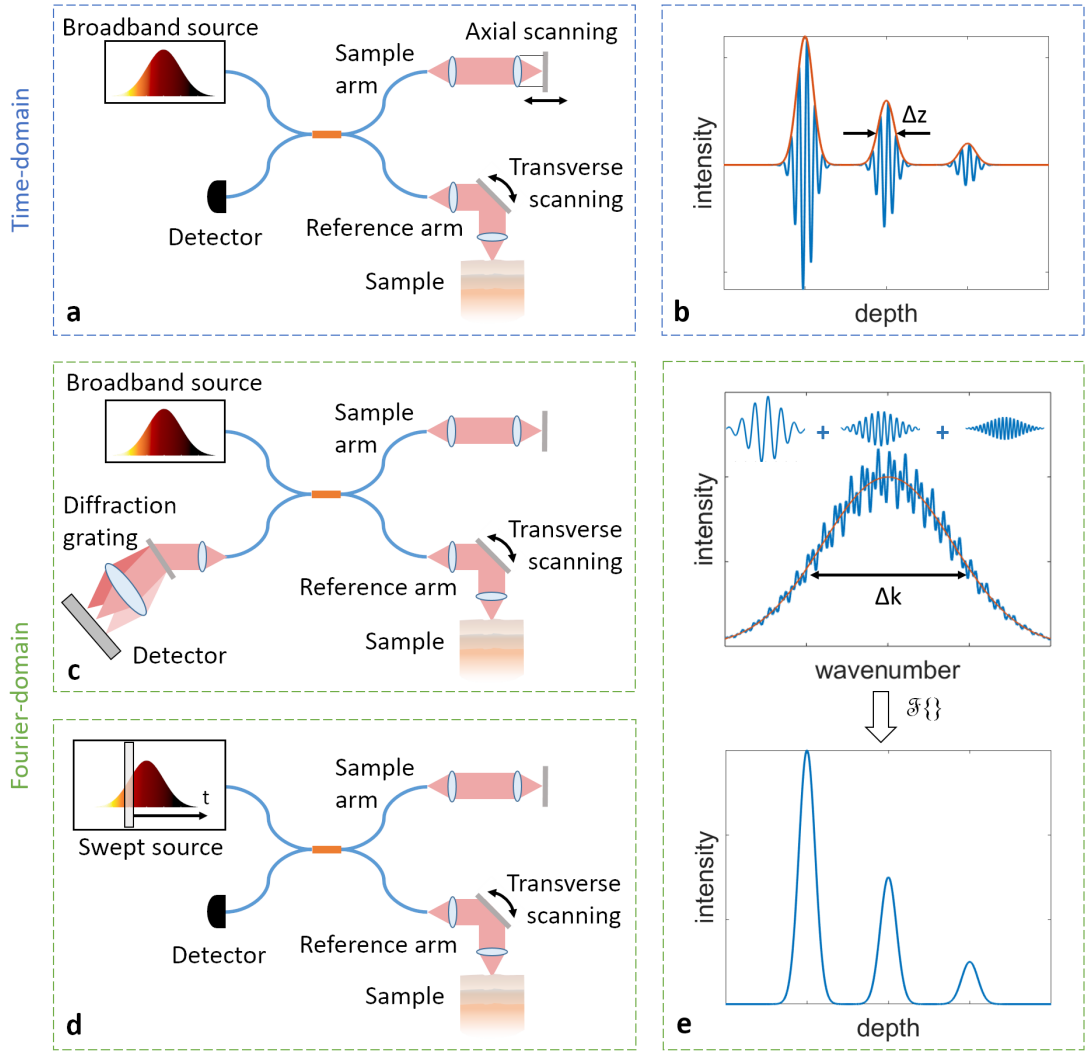

Figure 2.8: Optical coherence tomography (OCT) system working in the time-domain (a). Light from a broadband source with short coherence length allows coherence gating of the optical signal backscattered from the sample. The axial profile of the sample is uncovered from the envelope of the measured signal (b) while the reference arm path of a Michelson interferometer is scanned. The axial resolution Δz is given by the coherence length. In spectral-domain OCT, light of a broadband source is used and the combined signal from the two arms is dispersed in a spectrometer setup (c). In swept-source OCT the incident light is swept across a wide range of wavelengths, and each wavelength of the interference signal is separately detected (d). In both cases the axial profile is recovered by Fourier analysis of the spectrum formed by the fixed length reference arm and the sample (e). The coherence length (and hence the axial resolution) is ruled by the bandwidth of the spectrum (Δk).

and Fujimoto (2015, chapter 5). Nonetheless, the connection between the two can be intuitively understood relying on the Wiener-Khinchin theorem. Low-coherence interferometry reveals the cross-correlation between the sample and reference arm electric fields. The Wiener-Khinchin theorem states that under certain circumstances the autocorrelation function of a signal and its power spectral density are Fourier transform pairs:

$$R_{ii}(\tau) = \int_{-\infty}^{\infty} S(f) e^{-2\pi i f \tau} df \tag{2.10}$$

The inverse Fourier transform of the measured power spectral density ($S(f)$) is, therefore, equivalent to the autocorrelation of the signal at the interferometer output ($R_{ii}(\tau)$). The interferometer output is the sum of the reference and sample arm electric fields. Indeed, this involves the autocorrelation of the reference field (a Dirac delta function), the autocorrelation of the sample field (usually negligible with a much stronger reference arm signal), and the cross-correlation between the fields of the reference and sample arms (see Fig. 2.10).

A slightly inaccurate but more picturesque description can be given by considering the following. The source provides broadband light, from which each wavelength shall be considered individually. Consider a simple reflection from the sample, with a path length different than that of the reference arm. Light of a certain wavelength travels different path lengths in the two arms, therefore, after reflection, the two beams arrive back to the beamsplitter with different phases. For a certain wavelength component, the path difference might cause matching phases, and the intensity of this component is increased through constructive interference. A slightly longer wavelength component shall have slightly mismatched phase difference over the same path difference, and a somewhat lower increase in intensity accordingly. For longer and longer wavelength components, at some point the phase difference over the path difference will be integer multiple of 180°, and the component will suffer destructive interference. Consequently, the intensity of interference spectrum will show a sinusoidal modulation. It is also clear from this picture why a longer path difference causes higher frequency of spectral intensity modulation. For this, consider two slightly different wavelengths, and the interference in the beamsplitter in both cases. The phase difference over a certain path difference in the two cases will be different but close, if the path length difference is short. The phase difference builds up when the path difference is long, and will be further away for components with similar wavelengths. The effect is that at long path differences the modulation of the interference spectrum changes more abruptly, i.e. with higher frequency.

2.3.2 Axial resolution, transversal resolution and depth of focus

The axial resolution (Δz) is generally defined as half of the coherence length (l_c) (Fercher, 1996). The latter is the FWHM of the source signal's autocorrelation function, also known as the coherence function (Goodman, 2000, chapter 5). Given a Gaussian-shaped source spectrum with λ_0 centre wavelength and $\Delta\lambda$ FWHM bandwidth, the axial resolution is given as (Akcay *et al.*, 2002)

$$\Delta z = \frac{l_c}{2} = \frac{2\ln 2}{\pi}\frac{\lambda_0^2}{\Delta\lambda} \tag{2.11}$$

Based on this equation, the axial resolution in function of the bandwidth around certain centre wavelengths is depicted in Fig. 2.9.

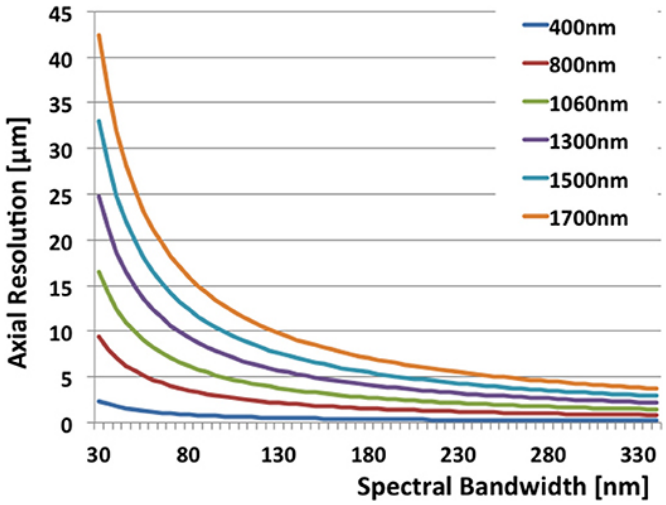

Figure 2.9: The achievable axial resolution in function of the source bandwidth at different centre wavelengths. Figure from Leitgeb and Baumann (2018).

In OCT, the axial resolution is independent from the lateral (transversal) resolution. While the former is determined by the coherence length of the source, the latter is governed by the sample arm optics and the focusing lens. The sample arm optics with fibre optic delivery is usually modelled as a confocal microscope with infinitesimally small pinhole aperture (Gu *et al.*, 1991). With a Gaussian spatial intensity profile of the illumination, the focused spot has a Gaussian intensity profile as well (Lexer *et al.*, 1999). The transversal resolution is either described by the width of the Gaussian shaped intensity at $1/e^2$ height ($\Delta x_{1/e^2}$ in equation 2.12) (Leitgeb *et al.*, 2006), or as FWHM (Δx_{FWHM} in equation 2.13) (Woolliams *et al.*, 2010).

$$\Delta x_{1/e^2} = \frac{4\lambda}{\pi}\frac{f}{d} \tag{2.12}$$

$$\Delta x_{FWHM} = \sqrt{\frac{\ln(2)}{2}}\Delta x_{1/e^2} \approx 0.59 \cdot \Delta x_{1/e^2} \approx 0.37\frac{\lambda}{NA} \tag{2.13}$$

Here λ is the (centre) wavelength, d is the width of the collimated beam falling to the focusing lens, f and NA are the focal length and numerical aperture of the focusing lens. Therefore, lenses with high numerical aperture (NA) are advantageous for fine transversal resolution. On the other hand, they negatively affect the depth of field where the probe beam is focused, often referred to as axial field of view (FOV) or depth of focus (DOF). A commonly used definition of the DOF in confocal microscopy is the FWHM of the detected intensity in function of the position of a planar reflector along the optical axis (equation 2.14) (Drexler and Fujimoto, 2015, chapter 2). With a small-angle approximation, this definition approaches a simpler expression (Sheppard, 1988), where the DOF is defined by the confocal parameter b, which is twice the Rayleigh length (equation 2.15) (Fujimoto *et al.*, 2000). In the case of Gaussian beams, the Rayleigh length (z_R) denotes the distance along the propagation of the beam at which the cross section of the beam is twice of that at the point of focus (i.e. the beam waist $\Delta x_{1/e^2}$).

$$dof = \frac{0.221 \cdot \lambda}{\sin^2[\frac{\arcsin(NA)}{2}]} \tag{2.14}$$

$$b = 2z_R = \frac{\pi \Delta x_{1/e^2}^2}{2\lambda} \tag{2.15}$$

In typical OCT applications, the focusing lens is of low NA, so that the DOF is longer than the coherence length. Consequently, the DOF does not affect the axial resolution, but limits the transversal resolution.

More accurate estimation of the axial and transversal resolution can be performed using resolution test artefacts and approximation of the system's location dependent point spread function (PSF) (Tomlins *et al.*, 2011). This method not only accounts for the effect of aberrations of the optical components and non-Gaussian source spectra, but also improves the resolution through deconvolution with the PSF (Woolliams *et al.*, 2010). It should be pointed out, that resolution is usually analysed using linearly scaled intensity values. Since OCT images are displayed over a logarithmic scale, substantially different intensity values are moved 'closer' to each other. Thus, the separation of two features that are to be resolved with the same fraction of intensity dip

between them (with respect to the peak intensity) requires a larger distance between the features (Woolliams and Tomlins, 2011). The multiplicative factor increasing the resolution in the logarithmic domain also depends on the dynamic range of the image (defined in chapter 5).

2.3.3 Light sources and detectors

Properties of the light source are particularly important in OCT. The centre wavelength and the bandwidth determines the achievable axial resolution (see Fig. 2.9). The centre wavelength needs to be in the NIR for sufficient penetration of light into tissue. At wavelengths closer to the visible range, much of the light is absorbed by melanin and haemoglobin in the tissue. Close to 2 µm, water becomes a dominant absorber. Increasing the centre wavelength of light decreases scattering in most biological tissues and allows deeper penetration, but compromises the depth resolution (Sharma *et al.*, 2008; Kodach *et al.*, 2010).

High power, broadband light with tens of nanometres bandwidth in the NIR range is available through relatively inexpensive superluminescent diode (SLD) sources (Drexler and Fujimoto, 2015, chapter 10). High performance lasers with femtosecond pulses and extremely broadband light (hundreds of nanometres bandwidth), and photonic crystal fibres are used for ultrahigh resolution (1–2 µm) imaging (Drexler *et al.*, 2015). The widespread industrial application of such sources is limited, due to the complexity and high cost inferred by the light source. Supercontinuum laser sources have been used more and more frequently, owing to their extremely wide spectral range in the visible and NIR region (Labruyère *et al.*, 2012). Early SS OCT systems applied lasers with semiconductor amplifiers and tunable filters (Huber *et al.*, 2005). The introduction of Fourier-domain mode locking (FDML) (Huber *et al.*, 2006) and tunable vertical-cavity surface emitting lasers (VECSELs) (Chang-Hasnain, 2000) has brought significant advancements in imaging speed and imaging range for SS OCT systems.

The traditional sensors in OCT have been CCD and CMOS detectors. The behaviour of these sensors in OCT is well understood and documented in literature. Mohan *et al.* (2008) demonstrated SPC detection in TD OCT systems using superconducting single-photon detectors (SSPDs) and SPADs. The large bandwidth and smooth spectral response of SSPDs were shown to increase the axial resolution, while SPC was shown to increase sensitivity, enabling nearly shot-noise-limited performance even with weak light sources. Muller and Fraser (2009) demonstrated time-gated detection of the optical signal for OCT images with improved contrast. Their optical system did not utilise special detectors, but did require the use of a nonlinear optical crystal and a high-end ultrafast laser. SPAD line sensors offer time-resolved detection with no readout noise. No study has focused on manipulation of the detection of scattered

light, based on its time-of-flight in an OCT setup, using time-resolved single-photon detection. This capability of the Ra sensors lead to ideas which are explored in Chapter 6.

2.3.4 Limitations of spectral-domain optical coherence tomography

SD OCT is highly advantageous compared to TD OCT. However, the spectral detection of SD OCT infers several artefacts which are not present in time-domain systems. This thesis places great emphasis on SD OCT, therefore its main limitations are presented.

Mirror artefact

The Fourier transform is Hermitian symmetric. In standard FD OCT, the depth profile is calculated as the magnitude of the inverse Fourier transform of the real-valued interference spectrum. Due to the Hermitian symmetry, the result is a signal that is symmetrical to zero position, which is set by the reference arm path length (Fig. 2.10). In effect, FD OCT images are folded around zero depth when features of the sample cross the reference path length. Mirror artefacts impose challenges when automatic segmentation and clinical diagnosis is to be made based on the tomography images (Ho *et al.*, 2010). The problem is not only that standard FD OCT cannot differentiate between positive and negative path differences, but also that the number of sampling points of the depth profile is half of the number of the interference spectrum sampling points, compromising the depth resolution.

Sampling effect

In SD OCT, the interference signal is recorded at discrete optical wavelengths (and frequencies), determined by the pixels of the detector array. Since the interference spectrum is sampled, its transform, the depth profile is also sampled at discrete depths. The sampling interval in the resolved depth profile depends on the detected spectral range. For this, consider a linearly spaced sampling of the interferogram in the k-domain (optical frequencies) on M pixels, with an interval of $\delta_s k$. The number of pixels determines the total detected spectral range $\Delta k = M \cdot \delta_s k$, which in turn determines the sampling interval in z-domain (resolved depth) according to $\delta_s z = \frac{2\pi}{\Delta k}$. Note, that the distance between sampling points in the real depth profile ($\delta_s \hat{z}$) is half of this, due to round-travel of light in the interferometer arms (see equation 2.16).

$$\delta_s \hat{z} = \frac{2\pi}{2\Delta k} \tag{2.16}$$

The total depth $\pm\hat{z}_{max}$ in the $\hat{z}$-domain 'seen' by the sensor is set by the number of pixels (M), as described by equation 2.17.

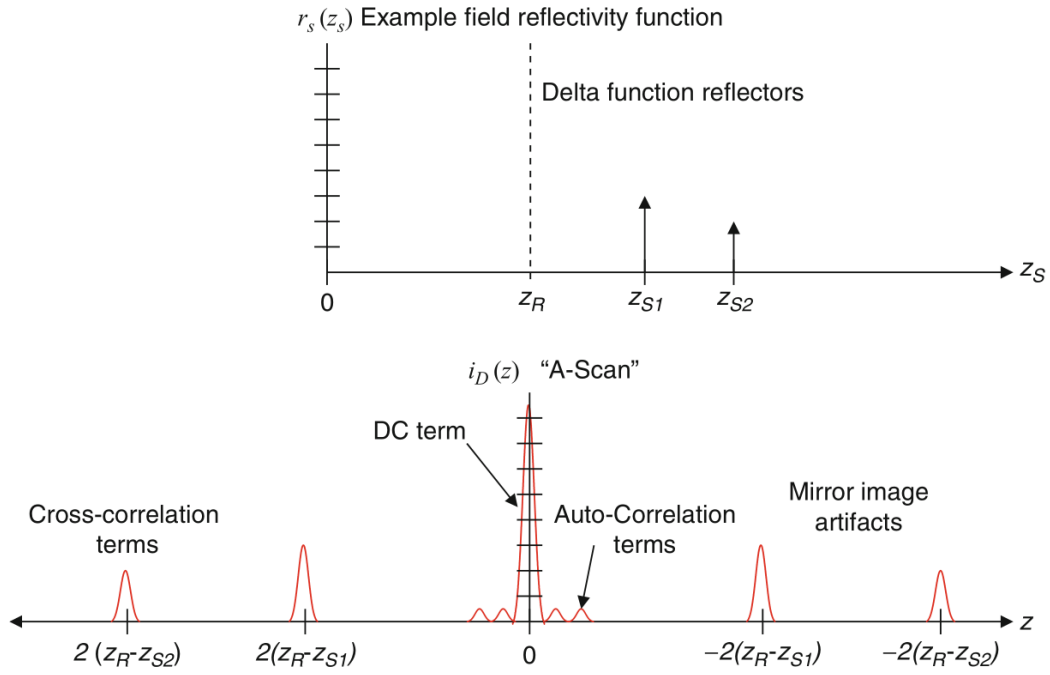

Figure 2.10: Illustration of a sample arrangement (top) and the respective signal, calculated through Fourier transformation of the interference spectrum (bottom). The sample arrangement consists of two Dirac-delta reflectors at positions z_{S1} and z_{S2} along the depth of the sample (z_S), and a reference mirror at position z_R. The Fourier transform of the interference spectrum reveals the autocorrelation of the combined sample and reference arm signals. This involves the cross-correlation between the signals of the reference arm and the sample arm reflectors, at depths belonging to their path differences. The factor 2 in the reflector positions is a consequence of round-travel of light in the optical setup. The A-scan also shows the autocorrelation of the reference signal (DC term) and the autocorrelation of the sample signal around zero depth. Mirror image of the A-scan is due to Fourier-transformation of a real-valued spectrum measurement. Figure from Drexler and Fujimoto (2015).

$$\pm\,\hat{z}_{max} = \pm\frac{M\cdot\delta_s\hat{z}}{2} = \pm\frac{\pi}{2\cdot\delta_s k} \tag{2.17}$$

Plus or minus signs of the total depth ($\pm\hat{z}_{max}$) belong to positive and negative spatial frequencies in the discrete Fourier transform of the real-valued interference spectrum, i.e. negative and positive path length differences with respect to the reference arm path length. Reflections from negative and positive path differences cannot be distinguished when real-valued spectra are detected, as discussed above.

Sensitivity falloff

The optics for dispersing the interference signal and the sensor detecting its spectrum have imperfections, negatively affecting the derived reflectivity profile. The effect is a depth dependent decrease in intensity, referred to as sensitivity falloff. The finite spectral resolution introduced by the spectrometer optics is usually modelled by a narrow Gaussian function (Häusler and Lindner, 1998) with a FWHM of $\delta_r k$, written as

$$R(k) = exp\Big[- \frac{4 \ln 2 \cdot k^2}{\delta_r k^2} \Big]$$

The ideal spectral response ($I(k)$) is convolved with the finite resolution of the spectrograph ($I(k) * R(k)$), and thus the optical signal falling to the sensor can be described as the real depth profile ($I(\hat{z})$) multiplied by a Gaussian decay (see equation 2.18). The effect of finite spectral resolution can be measured through recording a series of interference spectra with different time delays (path differences) (Dorrer *et al.*, 2000).

The recorded interference spectrum is also affected by the sensor pixels, as these have a rectangular aperture. The effect can be described as a convolution with a rectangular function (Leitgeb *et al.*, 2003). The sampling and the detector footprint together is written as a multiplication with a sinc squared function in the $\hat{z}$-domain (Yun *et al.*, 2003). The detected spectrum affected by the finite spectral resolution of both the spectrometer optics and the sensor is given in the $\hat{z}$-domain ($I_d(\hat{z})$) as

$$I_d(\hat{z}) = I(\hat{z}) \cdot exp\Big[- \frac{\hat{z}^2 \cdot \delta_r k^2}{4 \ln 2} \Big] \cdot sinc^2(\hat{z} \cdot \delta_s k) \tag{2.18}$$

with the pixel width $\delta_s k$, and FWHM optical spectral resolution $\delta_r k$.

2.3.5 Data processing

To retrieve the depth profile from the measured spectral interferogram, several processing steps have to be carried out, including the inverse Fourier transformation.

Firstly, background subtraction is performed on the recorded spectra. The background line is measured by the reference arm enabled only, and subtracted from each interference spectrum recorded during imaging. The differential measurement rejects the fixed pattern noise of the light source and the detector, and transforms the dark noise of the detector to have a distribution with zero mean. Background subtraction removes the DC component of the measured data before Fourier transformation, and hence removes the corresponding peak from the A-scans located at zero depth.

Fourier transformation connects physical distances (z-space) and wavenumbers (k-space). To perform transformation on the sampled spectrum, the fast Fourier trans-

form (FFT) is usually applied, which requires an equidistant sampling. In most systems, the diffraction gratings and sensors provide approximately even sampling in wavelengths, therefore the recorded spectra have to be resampled before FFT calculation. Effectively, this is done by defining the new sampling points based on the relation of wavelengths and wavenumbers ($k = \frac{2\pi}{\lambda}$). The interference intensity is estimated over these points using the recorded interference spectrum and spline interpolation.

Additionally, the amount of dispersion of light in the two arms is often different, due to travelling through different materials (e.g. tissue). Dispersion can be balanced prior to the inverse Fourier transformation by means of hardware (Hitzenberger *et al.*, 1999), or software (Wojtkowski *et al.*, 2004).

The intensity of backreflected and backscattered light versus depth is calculated from the complex-valued Fourier transformed data. For this, the absolute value of the Fourier transform is squared. With Hermitian symmetric transforms, half of the transformed data is discarded, being redundant (mirror artefact). In the final step, the transformed data is rescaled to a logarithmic scale (decibels), since the calculated values range over several magnitudes.

2.4 Conclusion

Three spectroscopic techniques have been described: fluorescence, Raman and OCT. This thesis analyses each via deployment of CMOS SPADs for massively parallelised time-resolved studies. For fluorescence, it is the fast fluorescence decays that can be measured spectrally (Krstajić *et al.*, 2015; Kufcsák *et al.*, 2017). In Raman, time-resolved features of detectors allow for separation of fluorescence and Raman signals (Ehrlich *et al.*, 2017b; Kekkonen *et al.*, 2018) opening the possibility of either increasing the SNR of Raman spectra detected or increasing the speed of acquisition. Lastly, this thesis introduces the first ever implementation of time-resolved SD OCT using CMOS SPADs. The combination of interferometry, single photon counting and time-resolved electronics opens up new possibilities, as described in Chapter 5 and 6. The unifying themes in this thesis are the time-resolved approach and its implementation via CMOS SPADs. The last chapter, Chapter 7, with its conclusion and appendices gives a good overview of the way forward and future experiments.

Chapter 3

Characterisation of CMOS SPAD Line Sensors

An ideal sensor operating in photon counting mode detects every single photon hitting the sensor, and registers only the incident photons, only once. Real devices, however, have limitations regarding the range of photon energies they can detect and a finite probability of detection. Also, an output signal of the device does not necessarily account for a detected photon, as multiple counts can be registered for a single event or outputs can be generated even in the absence of incident photons. Therefore, the sensor's performance has to be evaluated and various parameters of it have to be measured before deploying it in any application. In this thesis two complementary metal-oxide-semiconductor (CMOS) single-photon avalanche diode (SPAD) line sensors were used, Ra-I and its successor, Ra-II. Both of these sensors were designed by Dr. Ahmet Erdogan and Dr. Richard Walker at the CMOS Sensors and Systems Group of the University of Edinburgh. Ra-I and Ra-II were developed as part of the Proteus project (`http://proteus.ac.uk`), primarily targeting time-resolved fluorescence spectroscopy. In chapter 4, Ra-I was deployed, as the first available sensor. Ra-II became ready for use after the characterisation and application of Ra-I (including firmware and software work). Following studies (described in chapters 5 and 6) used the improved, Ra-II sensor. In this chapter, frequently used measures are discussed first to describe CMOS SPAD sensors. Following this, the methods for measuring the most important parameters of the Ra-I and Ra-II sensors are presented, such as noise, sensitivity and timing performance, and the respective results are discussed. General characterisation of Ra-II was carried out by Dr. Ahmet Erdogan. Some of these characterisation results are also presented in this chapter, and acknowledged appropriately.

3.1 Figures of merit

Different fields and companies use a variety of different measures to express the effectiveness of a sensor and the goodness of its characteristics such as noise level or the weakest and strongest detectable signals. Terms like sensitivity or efficiency are often confusing and not directly comparable when used for different sensor technologies. This section aims to describe and define the measures which are most often used when evaluating SPAD based sensors. Most of these measures are determined at the design level, while some of them are accessible for the end user to tune, by carefully setting the operating conditions and parameters of the sensor and its SPADs (e.g. bias voltage). These parameters often favour a certain property while infer an undesired condition regarding another property of photon detection. Therefore, it is even more important to have an insight on the characteristics of SPAD sensors and how they are affected by sensor parameters.

3.1.1 Detection characteristics

Photon detection probability, photon detection efficiency and fill factor

Photon detection probability (PDP) is a measure of the sensor's sensitivity to incident photons impinging on the active area of the sensor. PDP can be expressed as the product of two terms, the probability of charge carrier generation by an incident photon (quantum efficiency (QE)), and the probability of a photo-generated carrier inducing an avalanche breakdown (Fishburn, 2012).

$$PDP = QE \cdot P_{avalanche}$$

Photon detection efficiency (PDE) is usually used to describe sensitivity, also taking into account the restricted sensitive area of the sensor. Limitation on the sensitive area is caused by the SPAD structure, and ancillary circuitry on the silicon surface for operating the SPADs (e.g. timing and readout electronics). PDE can be expressed as the product of PDP and the fill factor (FF) of the sensor, where FF is the ratio of the area sensitive to light and the total area (Charbon, 2014).

$$PDE = PDP \cdot FF$$

PDP and PDE are functions of the wavelength of the incident light and also depend on the bias voltage of the SPADs. Typical PDP values are in the range of 20–40 % to visible light (Charbon, 2014), with reports on peak PDP over 40 % (Webster *et al.*, 2012a; Veerappan and Charbon, 2015), achieved with optimized SPAD structures.

Increased sensitivity through increased FF has also been demonstrated using back-side illumination (BSI) and applying 3D stacking processes (Pavia *et al.*, 2015; Abbas *et al.*, 2016). With a given FF, more efficient photon collection can be achieved with microlens arrays placed over the sensor (Pavia *et al.*, 2014; Gyongy *et al.*, 2018).

Dark count rate

The main noise contributor of a CMOS SPAD sensor is the noise associated with dark counts. Dark counts are avalanche breakdowns caused by charge carriers reaching an excited state through non-photonic interactions, such as thermal generation, background radiation and tunnelling. The avalanche breakdowns due to these charge carriers manifest as photon detections, even when the sensor is not exposed to light (hence the name dark counts). The rate of dark counts limits the smallest detectable optical signal, which is of particular importance in low light level scenarios. Dark count rate (DCR) describes the average number of dark counts in a given time interval, most often expressed in units of counts per second (CPS) or hertz (Hz).

DCR depends on many factors such as the size and shape of the SPADs or the manufacturing process of the sensor (Xu *et al.*, 2017). DCR also changes with the bias voltage of a SPAD, since increasing bias voltages increase the electric field across the device and the probability of an avalanche breakdown. Several methods have been advised in order to mitigate DCR of a sensor after design and manufacturing. A common technique to reduce the contribution of thermally generated carriers to DCR is to cool down the sensor (Rech *et al.*, 2004; Villa *et al.*, 2012; Buttafava *et al.*, 2014; Gyongy *et al.*, 2016a). Turning off detection elements of high DCR could also be beneficial (Finlayson *et al.*, 2018). Time-resolved sensors have the capability of limiting the detection in time to a region when a signal of interest is expected (Nissinen *et al.*, 2011; Gyongy *et al.*, 2016a). Post processing algorithms are also widely applied to reduce noise (Li *et al.*, 2009; Maruyama and Charbon, 2011; Gariepy *et al.*, 2015; Gyongy *et al.*, 2016b).

Deadtime

Following the buildup of an avalanche breakdown, the avalanche current has to be quenched and the SPAD restored to a sensitive state so it is ready again for photon detection. Quenching is performed by lowering the bias voltage across the device, which is followed by recharging the SPAD (Fig. 3.1). The duration of quenching (quenching time) and recharging (reset time) during which the SPAD is incapable of detecting light is referred to as deadtime, and it is usually in the range of 10–100 ns (Bronzi *et al.*, 2016). Deadtime restricts the number of photons that can be detected during a certain

time interval, and hence limits the detection bandwidth and the light level (or incoming photon rate) that can be detected without saturation of the detector.

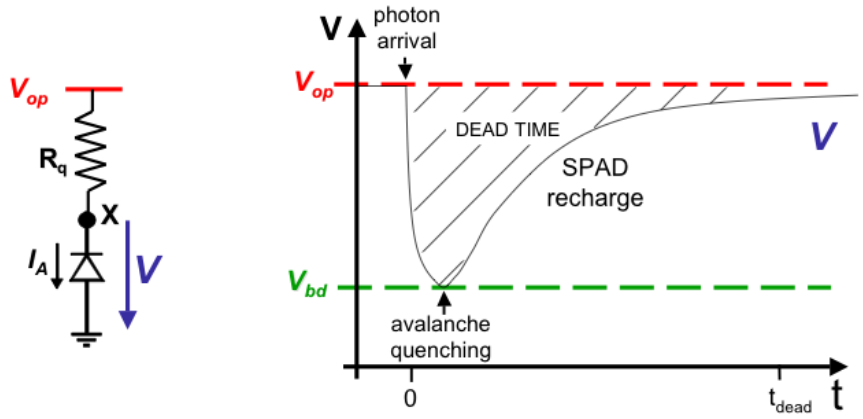

Figure 3.1: Quenching and recharging. Figure from Charbon *et al.* (2006)

The length of the deadtime greatly depends on the quenching method (Gallivanoni *et al.*, 2010). In the simplest form, quenching and recharging is carried out passively by connecting a ballast resistor in series with the SPAD. The avalanche current causes a voltage drop over the resistor, which decreases the voltage over the SPAD and suppresses the breakdown current. As the current is quenched, the voltage over the SPAD increases again to its original value, making the SPAD sensitive again. Passive quenching has the drawbacks of typically long and not well defined deadtimes, however it is simple and favours FF. In active quenching circuits, the onset of the avalanche is sensed through a low impedance, followed by active modulation of the bias voltage using fast switches. While this technique infers more complex design and additional circuitry lowering the FF, it provides shorter deadtimes and a higher flexibility of quenching. Spatial and temporal compression techniques (Braga *et al.*, 2014) rely on the fact that the counting electronics can be shared by neighbouring detection elements where the probability of simultaneous photon detection is low. Multiple SPAD outputs connected together by this way overcome the problem of insensitivity during deadtime of a single SPAD, and increases the level at which the sensor saturates.

Afterpulsing and crosstalk

Afterpulsing refers to the phenomenon when multiple output pulses are registered after the avalanche breakdown, generated by a photon detection. Afterpulses are generally explained as avalanche breakdowns, caused by the release of a charge carrier trapped in the semiconductor structure from a previous avalanche (Cova *et al.*, 1991). The quenching circuit (Ziarkash *et al.*, 2018) and photon detection at the end of the deadtime (twilight photons) can also be responsible for unwanted re-triggering

of the device (Polyakov and Migdall, 2007). When exploring the temporal distribution of photon arrivals, afterpulsing introduces an appearance of time correlation between otherwise uncorrelated incident photons. This time correlation can be used to measure the tendency of a device for afterpulsing, which is expressed quantitatively by the afterpulsing probability (Para, 2015). The presence of these spurious pulses not only distorts the measurement of incident photons, but also increases the time during which the device is incapable of detection, limiting the count rate that the sensor is capable of detecting. Afterpulsing probability sharply increases with the bias voltage (Stipčević, 2009), and incoming photon count rate (Ware *et al.*, 2007) due to higher probability of detecting twilight photons. Afterpulsing probability also increases with lower temperatures, as the time-constant of releasing trapped carriers increases above the deadtime of the device. Design considerations can limit afterpulsing probability by optimizing the quenching circuit and deadtime (Bronzi *et al.*, 2013). Time-gated operation of SPADs has also been shown to reduce the afterpulsing probability (Palubiak *et al.*, 2015).

Crosstalk is a similar effect to afterpulsing in the meaning that it describes spurious, non-photonic output pulses of the device, induced by previous photon detections. However, in the case of crosstalk the false counts appear on neighbouring SPADs to the one detecting a true event. The origin of crosstalk could be electric, due to capacitive coupling between neighbouring devices. The cause of crosstalk could also be optical, when photons from secondary emission are absorbed and induce an avalanche breakdown. Photons of secondary emission are generated during the relaxation and recombination of charge carriers from a previous avalanche process. The active region of SPADs is surrounded by a guard ring. The primary aim of the guard ring is to prevent premature edge breakdown of the device (breakdown due to the excessively high electrical field at the edges of the junction and implant boundaries). However, the guard ring also plays an important role in isolating neighbouring SPADs and preventing crosstalk (Xu *et al.*, 2014). Crosstalk depends not only on the wavelength and intensity of the incident light, but also on structural and geometrical properties of the sensor such as the guard ring, distance between SPADs, and thickness of the wafer, and hence can be tackled during the design phase of the sensor (Rech *et al.*, 2008; Ficorella *et al.*, 2016).

3.1.2 Timing characteristics

Timing jitter and instrument response function

The timing jitter of a SPAD refers to the variation of the time elapsed between the absorption of a photon and detecting the leading edge of its corresponding SPAD output signal. Various factors affecting jitter include intrinsic properties of the device and the ancillary circuits detecting the avalanche breakdown. Intrinsic contributions to jitter are present at several stages of the avalanche buildup: at charge carrier generation (e.g. statistical fluctuations of the position of charge carrier generation and the time of their diffusion to the depleted region), or at the avalanche multiplication (e.g. vertical spread of the avalanche) (Spinelli and Lacaita, 1997). Most factors of jitter are, therefore, determined at the design stage. The jitter from an end user point of view can be optimized by selecting the right bias voltage. The jitter decreases with increasing bias voltages, as the increased electrical field across the device leads to faster buildup of the avalanche current. Jitter is usually expressed in function of the wavelength of incident photons, and its typical values range from few times ten to hundreds of picoseconds (Bronzi *et al.*, 2016).

Jitter limits the timing performance of the sensor, however in many sensing and imaging applications the whole system's timing uncertainty is of importance instead of that of a single detection device. In these cases, the overall timing variation of the system has to be measured, for which the instrument response function (IRF) is a more accurate term to be used. Beyond SPAD jitter, the IRF includes contributions from the surrounding circuitry of SPADs and the whole optical setup. Major sources of timing uncertainties on sensor level involve an accumulated jitter when a pixel consists of several SPADs connected together, and uncertainties of the timing circuits (time-to-digital converters (TDCs), time-to-analogue converters (TACs)). In the optical setup, the most important factors to consider are variations related to light pulses of the light source (e.g. variations in pulse width and period), and its corresponding electrical sync signal. Accurate estimation of the IRF can be carried out by using a highly scattering material as a sample (Lakowicz, 2006, Section 4.3). In many cases, especially where the timing uncertainties of the optical setup and the light source are negligible compared to the detector jitter, the jitter of the detector is measured only, but it is still referred to as the IRF.

Time-to-digital converter resolution

One key aspect of time-resolved applications is how accurately the arrival time of incident photons can be measured. Often in the case of SPAD based detectors (and with the Ra sensors in particular) detection of photons and retrieving information on their time of arrival are carried out in different steps. Namely, an electric pulse generated upon the avalanche breakdown (detection) is compared to an electric pulse belonging to a reference event, typically the sync signal of the pulsed light source (timing). Since timing is detached from photon detection in this sense, it is relevant to assess the precision that the timing circuits provide. Several different methods have been proposed for the comparison of two events (e.g. sync pulse and photon detection), and estimating the time interval between them with high accuracy (Kalisz, 2004). When analysing timing instruments which provide measurement results as digital data (TDCs), the most important parameters are the resolution of timing, (the smallest time difference that is expressed as a single digital unit), and the uniformity of conversion from time to digital data. For the latter, performance measures of data converters are used, such as integral nonlinearity (INL) and differential nonlinearity (DNL). Typical resolution of integrated TDCs is in the range of tens and hundreds of picoseconds (Palubiak and Deen, 2014). Less then 10 ps resolution was achieved with TDCs realised on a field-programmable gate array (FPGA) (Menninga *et al.*, 2012).

Delay resolution

In timing instruments, it is also often desired to add a variable delay to electrical signals (such as a sync signal from the light source). Independently from the applied method and the underlying structure of the delay system, one could generally think of this as a digital-to-time converter, where similar measures can be characterised as with TDCs, such as resolution, INL and DNL.

3.2 Ra sensors

3.2.1 Ra-I

The Ra-I sensor has two rows of 256 pixels. Both rows of pixels are formed of 4-4 SPADs in a column. The SPADs of one row are designed to have peak efficiency at the blue end of the visible spectrum (hence denoted as ’blue SPADs’ and ’blue pixels’). The structure of SPADs in the other row of pixels is targeted to have high detection efficiency for photons at the red end of the visible spectrum and near infrared (NIR) wavelengths (’red SPADs’ and ’red pixels’). Each red and blue pixel pair is connected to a TDC. The TDCs are independent from each other and provide massively parallel timestamping of photons. The sensor can work in three different operation modes.

Single-photon counting In single-photon counting (SPC) mode, each pixel is connected to an asynchronous counter of 27 bit depth. These counters are built up of 27 cascaded circuits with two states (flip-flops). A pulse of photon absorption changes the state of the first flip-flop, the output of which is used to initiate switching of the second flip-flop, and so on. With each photon, the pulse information ripples through the connected flip-flops (hence often called a ripple counter), to switch the whole structure to a different state, which can, therefore, count a total of 2^{27}-1 photons. The counters can be enabled/disabled for gated photon counting by a time-gate signal of user defined width and position. The bias voltage of SPADs is not modulated during gating, only the counters are masked, i.e. the SPADs are not turned off, but the counters are disabled for a certain time period, therefore the pulses from photon absorption are not counted. This method will be denoted as time-masking throughout this work, as opposed to time-gating which refers to turning the SPADs on/off. The user can set the time period when the counters are disabled/enabled with respect to an external trigger signal (e.g. from the pulsed laser of the optical system). The trigger pulse is delayed in two instances, one is used to enable the counters, and the other is used to disable them again. This way, the two delayed pulses define the rising and falling edges of the time-mask, and so the temporal position of the mask in which photons are to be counted. Delaying the pulses in time is carried out by sending the signal through a line of units, each causing a short time-lag (delay elements). The total delay applied to the pulses is set by selecting the number of delay elements they shall pass through. This is achieved by 128:1 multiplexers, selecting which of their 128 input should be connected to their output, based on how they are controlled (see Fig. 3.2). Separate multiplexers are used for defining the amount of delay elements for the rising and falling edges, hence these two delays can be defined independently. The rising and falling edges of the mask are routed for each of the 256 pixels by tree shaped wire structures (clock trees). This is to maintain that the pulses travel the same distance to each pixel, therefore the same additional delay adds to them due to propagation time, and that the mask is applied at the same time across the line of pixels.

Time-correlated single-photon counting In time-correlated single-photon counting (TCSPC) mode of the sensor, the first photon detection in an exposure cycle starts the time-to-digital converter of the pixel. The TDC is based on a gated ring oscillator (Richardson *et al.*, 2009). In general, a ring oscillator produces an oscillation between two voltages with a specific frequency, by inverting a logical one value to logical zero and then to a logical one again, and so on, in an endless loop. This oscillation is used for timing of events.

In practice, the TDCs of the Ra sensors employ a slightly different ring oscillator structure (Fig. 3.3). Here, 4 inverter elements are connected with the last one tied to

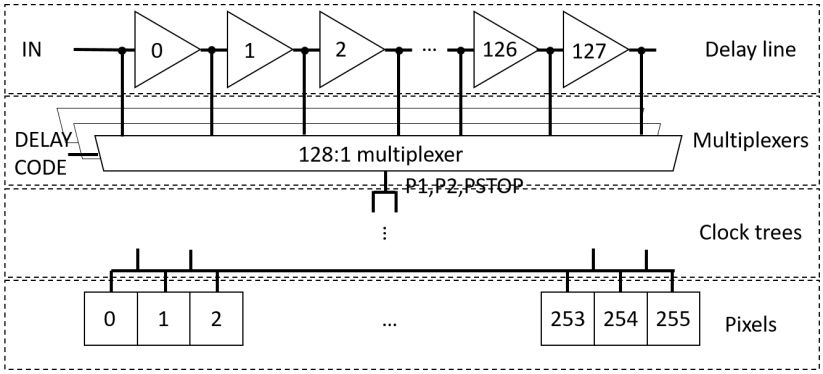

Figure 3.2: Delay generator architecture of Ra-I. A trigger pulse (IN) is delayed as passing through multiple delay elements (delay line), each causing a time-lag. Delay for the edges of the time-mask are set by 128:1 multiplexers. Depending on how they are controlled (through DELAY CODE), the multiplexers connect one of their inputs to their output, and hence set the amount of delay the trigger pulse shall go through before reaching the clock trees. The clock trees provide equal wire lengths to all pixels, so that the propagation time is the same for all signals, and the mask position is the same for all pixels.

the first, forming a loop. A consequence of this structure is that the ring oscillator has 8 different phases (or states). After start (with the START signal, provided at photon absorption), the ring oscillator iterates through the 8 phases, continuously. A ripple counter is increased with each loop. The ring oscillator is stopped by a STOP signal (usually provided by the laser sync output, as described in the previous chapter). At this point, the value of the counter gives a rough description of the time between the photon detection and the STOP signal. The state of the ring oscillator within the current loop can be determined with a coder unit, for fine timing information. In Ra-I, only the value of the counter is used for timing, and the remaining number of phases (not adding up for a turn of the counter) are neglected. Therefore, the TDC resolution is slightly lower than it would be with the state of the ring oscillator. However, due to the coarse resolution of the TDC, it is also possible to measure longer time periods (and analyse phosphorescence lifetimes in the µs and ms range). At the end of the exposure cycle, each pixel's counter value is latched to registers, waiting for readout.

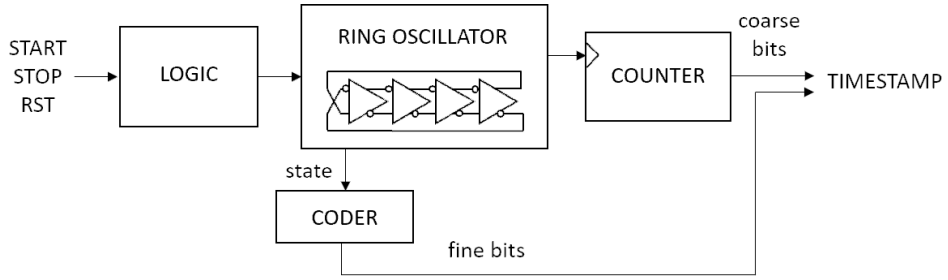

Figure 3.3: Simplified architecture of the TDCs of the Ra sensors. A logical block controls starting, stopping and resetting the ring oscillator. The ring oscillator increases a counter at each loop of 8 phases, providing a coarse timestamp. Finer timing information is provided by coding the phase of the ring oscillator when it was stopped.

'Centre of mass' mode The third operation mode is the 'centre of mass' mode (CMM), providing the average photon arrival time. In this mode of the sensor, up to one photon per laser pulse can be processed, providing higher photon processing rates than raw TCSPC. The first photon detected in a laser cycle (between two STOP pulses) starts the TDC, while the following STOP pulse freezes it, but does not reset it. As a result, the 27 bit counter sums the photon arrival times. A separate 16 bit counter counts the detected photon occurrences. At the end of an exposure cycle, the two numbers are divided off-chip and provide the centre of mass of photon arrival times. In most cases, when detecting exponential decays (e.g. in fluorescence spectroscopy), the centre of mass of photon arrival times is slightly different than the lifetime of the decay. Still, these two are correlated, and decays with different lifetimes have different centre of mass values as well. Since the centre of mass is almost instantly given by the sensor (as opposed to the more complex task of decay lifetime estimation), this operation mode is highly beneficial when the lifetime is of secondary importance and one only wants to differentiate between decays with different lifetimes. Both TCSPC and CMM can be operated in conjunction with the above described counter masking scheme.

The schematic architecture of a Ra-I pixel is depicted in figure 3.4. Each pixel receives the delayed pulses for the temporal mask edges (P1 and P2) and for stopping the TDC (PSTOP). A dedicated block (MASK LOGIC) takes care of masking the lines on which the pulses of photon absorption travel towards the counters, for the period defined by the time mask. 4 blue and 4 red SPADs are connected together with an OR gate, so that a photon absorption in any of the 4 SPADs is propagated further on the same line. A data path selector (multiplexer, controlled by the SPAD SELECT signal) selects which set of SPADs (red or blue) shall be used. If SPC mode is selected (by the SPC SELECT signal), the pulses of photon absorption are connected to the 27 bit counter, increasing it with each detected photon. In TCSPC and CMM modes, the photon pulses are sending the START signal to a block, which also receives the STOP

signal and controls the timing process (TDC LOGIC). This involves controlling of the ring oscillator and counting of photons on the 16 bit counter, in CMM mode. In the time-resolved modes (TCSPC and CMM), the ring oscillator is connected to the 27 bit counter, and increases it with every loop. In Ra-I, the state of the ring oscillator is not used for the timestamps. The additional write (WR) and reset (RST) signals take care of writing the counter values to data storage units (LATCH) and resetting the internal units of the architecture, respectively. The information provided by the counters of the pixels are summarised in Table 3.1, with the benefit and drawback of each mode highlighted.

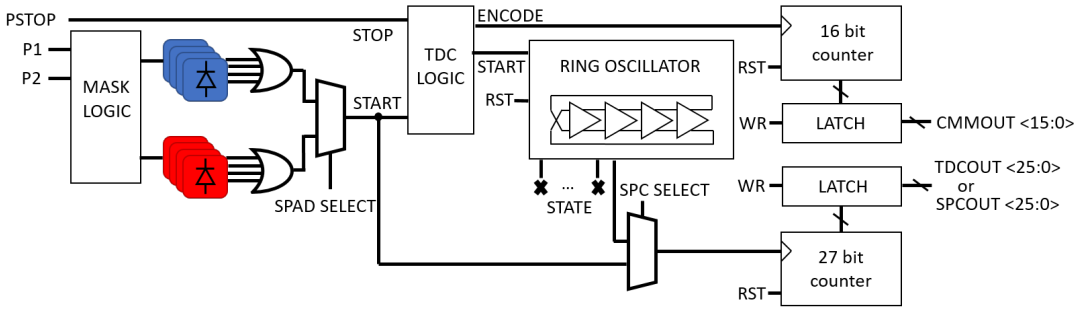

Figure 3.4: Pixel architecture of Ra-I, depicting the most important blocks and control signals used to apply temporal masking (MASK LOGIC block with signals P1 and P2), to select the set of four blue or red SPADs to be used (SPAD SELECT), to control the timing (TDC LOGIC block using the START signal by photon absorption and the STOP signal by the external PSTOP signal), and to increase the counters either at photon absorptions (in single-photon counting mode) or with each loop of the ring oscillator (in time-resolved mode, controlled by SPC SELECT).

For interfacing the Ra sensors, a custom printed circuit board (PCB) was designed by Dr. Ahmet Erdogan. This PCB also carries an extension board with an FPGA and USB2 connection. A firmware was written using the Verilog language, to control the sensor, read data from it (values of the 27 and 16 bit counters) and send the data to a host computer. The PCB and firmware are described in detail in the next chapter.

	SPC	TCSPC	CMM
27 bit counter	N	t_1	Σ_i t_i
16 bit counter		1	N
Benefit	High efficiency	Temporal information	Efficiency between SPC and TCSPC
Drawback	No temporal information	Low efficiency	Crude temporal information only

Table 3.1: Summary of different modes of Ra-I, showing the values of the counters and the benefits/drawbacks of each mode. Efficiency is defined as the maximum number of photons that can be processed during an exposure. N - number of photons, t_i - arrival time of photon *i*.

3.2.2 Ra-II

The design of Ra-II predates the characterisation and application of Ra-I, however it was intended to improve on many properties of Ra-I. These improvements include: fixing of design flaws of Ra-I, increasing the number of SPAD columns from 256 to 1024 and the number of SPADs per pixel from 4 to 8, increasing the number of timing circuits (TDCs) from 256 to 512 and increasing the TDC resolution. Additionally, new features were introduced, such as dividing/combining SPAD columns, double time-mask and time-gating feature (through switching the SPADs on/off), and on-chip histogram generation. On-chip histogram generation provides much higher photon processing efficiency than TCSPC, and yet records the entire time profile of optical phenomena, not just an average arrival time.

Ra-II has 1024×8 red and the same amount of blue SPADs, in two rows. 8 SPADs of a column are connected in an OR fashion. Each SPAD has its own enabling signal allowing to turn off SPADs with high DCR. Similarly to Ra-I, the Ra-II sensor can also work both in SPC, and time-resolved photon counting modes. There are two separate 20 bit photon counters for every two SPAD columns (counter A and B in Fig. 3.5) to assist SPC. When the two neighbouring columns are connected, photons falling to a 2×8 SPAD sized pixel are counted on both counters. To facilitate applications requiring higher spectral resolution, i.e. more pixels, the left and right columns of SPADs can be independently used, forming 1024 pixels, each owning a single column of 8 SPADs. In this case, each column is connected to one of the two counters.

Ra-II also provides means for gated/masked photon counting. Ring oscillator based delay generators are used to define 4 delayed pulses (with respect to a trigger event). These delay generators work similarly as the TDCs, except that in this case the required delay is preloaded, and the output pulse is generated when the timing structure has been running for the required time (Fig. 3.6). The four delayed pulses (P1 to P4) can be used to define two time masks (Fig. 3.7a), each masking one of the counters (left and right column counters in Fig. 3.5). In a different configuration, time-gating can be performed by lowering the bias voltage of the blue SPADs, and hence turning the SPAD entirely off. In this case, two of the four delay channels are used to define the time period for enabling blue SPADs. To allow time for the SPADs to turn on completely without breaking down due to photon impingement, a quench period is defined, using a third delay channel. Photons are detected only while the SPADs are enabled, but no quenching is carried out (Fig. 3.7b). Time-gated photon counting (i.e. turning the SPADs off) can be used with additional two time-masks (i.e. disabling photon counting), but the rising edge of the two time-masks has to match in this case, due to the limited number of available delay channels. Fig. 3.7 shows the use of the four delayed pulses (P1-P4) to define time-gates and/or time-masks.

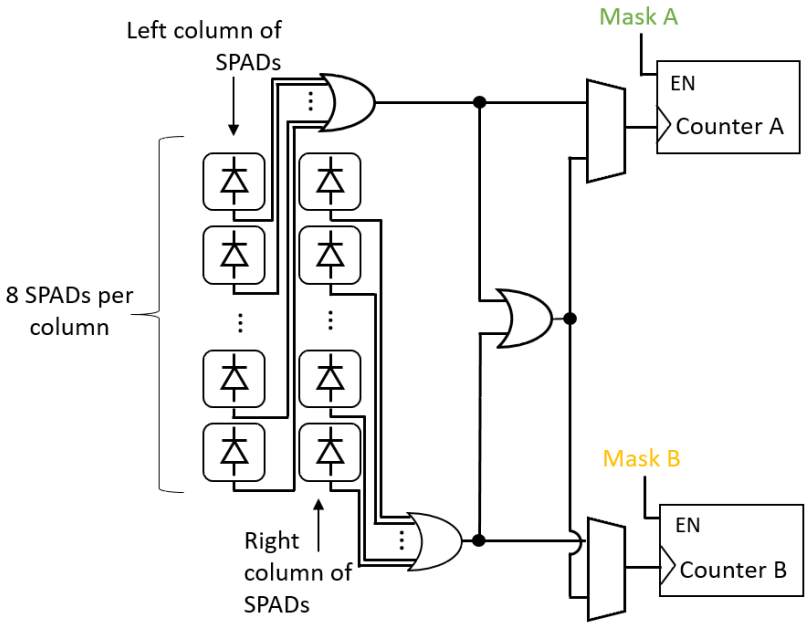

Figure 3.5: Architect enabling different pixel size configurations on Ra-II. The left and the right column of SPAD are both connected together through OR gates. The two columns are connected to separate counters (left column to counter A, right column to counter B) when the first input of the data selectors (multiplexers) is used. In this case the two columns work as separate pixels. The two columns are also connected together with an OR gate in the middle. If the second input of the multiplexers is selected, both counters are connected to this node, where the columns are joint. Two separate time-masks can be used at the same time in this case.

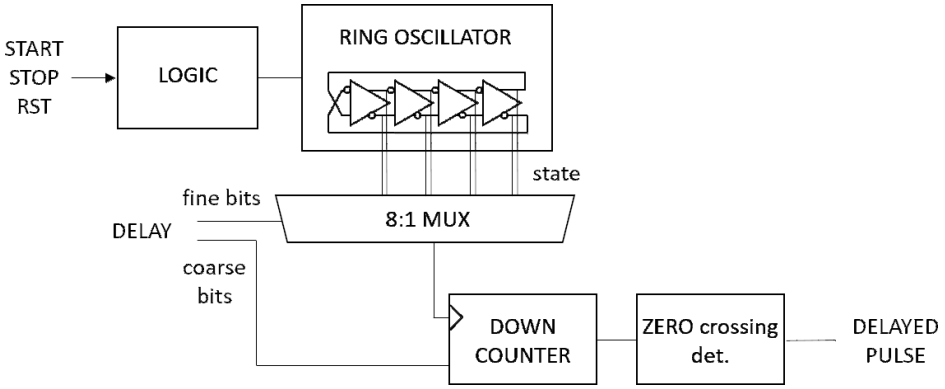

Figure 3.6: Ring oscillator based delay generator of Ra-II. Coarse bits of the required delay are used to load a counter. At each loop of the ring oscillator, the counter is decreased. Fine bits of the delay specifies at which phase of the ring oscillator the counter shall be stepped. This is achieved by connecting one of the internal nodes of the oscillator to the counter input, through a multiplexer (MUX). Finishing the countdown is marked when the counter crosses zero, at which point the output pulse is generated.

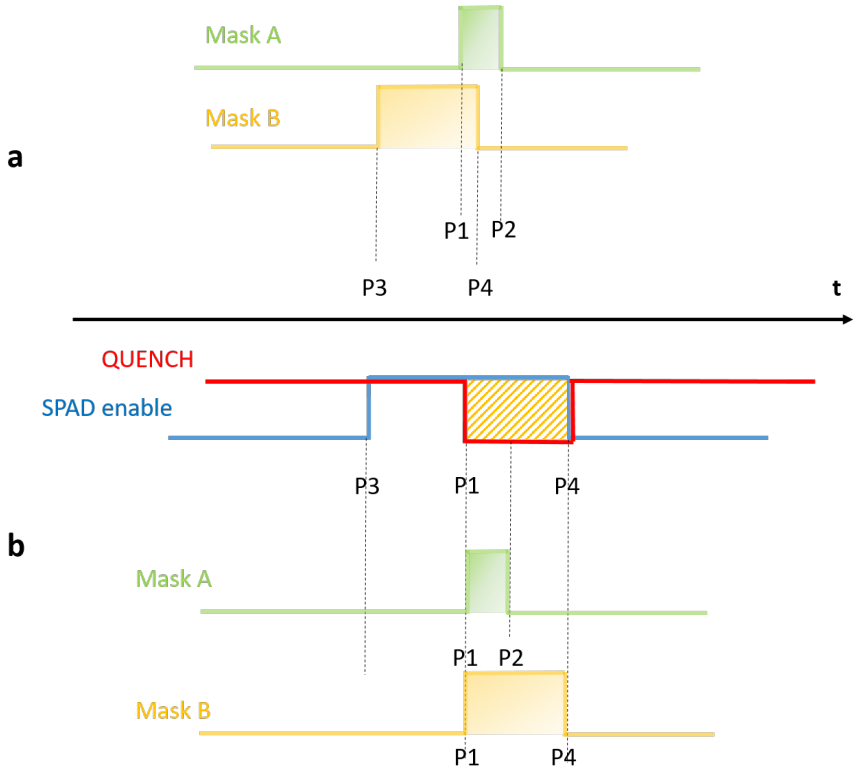

Figure 3.7: Time-gate and time-masks are defined by four delayed pulses (P1 to P4), generated by ring oscillator based, on-chip delay generators. Two independent time-masks can be used (a), or time-gating with additional masking (b).

Time-resolved photon detection is underpinned by 512 high resolution TDCs, using ring oscillators. TDCs are started upon the detection of photons and stopped when receiving a STOP signal (usually the sync signal from the light source), according to the stop-start operation principle. The 16 bit output of each TDC is built up from 3 bits, given by decoding the state of ring oscillators, and 13 bits coarse timing from a ripple counter, incremented at a certain state of the ring oscillator. The output of the TDCs can be used in two different ways, with respect to different time-resolved modes of the sensor. In TCSPC mode, the raw TDC timestamps are 'manually' latched to the sensor output, and the TDCs are reset at the end of an exposure cycle. Effectively, this is carried out by signals, controlled by the custom firmware running on the FPGA. To overcome the photon conversion-rate limitations usually associated with TCSPC operation, a different approach is also implemented on the Ra-II sensor, providing TCSPC histograms on-chip. In the on-chip histogram mode, several counters of the sensor are associated with histogram bins. One could deploy either a maximum of 16 counters, each with 20 bit depth, or up to 32 counters of 10 bit depth in each pixel, depending on the application requirements. TDCs in this mode are reset after every STOP pulse, and hence the photon conversion rate can go as high as the repetition

rate of the light source. The value of a pixel's TDC is used to address counters (time bins) within the pixel. When an incident photon is detected, it is registered on the counter which is enabled at the time of arrival. By selecting which TDC bits are used for enabling the counters (histogram bins), this approach provides a zooming capability over the histogram. Consequently, this approach offers a trade-off between time resolution of the bin and the total time interval, over which photons can be registered into the histogram. Both TCSPC and on-chip histogram modes can be further time-masked, by using the counter masking scheme (masking). Temporal sectioning in these modes can also be achieved by simply selecting certain time-bins of interest from the recorded histograms (windowing). Details of the architecture of Ra-II and related block diagrams are provided by Erdogan *et al.* (2019).

3.3 Methods for characterising the Ra sensors

The Ra-I sensor was characterised for DCR, IRF, TDC and delay line resolution. Deadtime, PDE and afterpulsing measurements were not carried out, either because of the complexity of measurements (as in the case of the PDE), or because of being of secondary importance. Given a valid detector model with the deadtime as parameter (see equations 5.11 and 5.12 in chapter 5), the deadtime can be estimated by fitting the model to measurements of the detected count rate in function of the incident count rate. Such measurement of the deadtime was not needed in the studies of this thesis, for the following reasons. As discussed, in TCSPC mode the average number of incident photons in an exposure cycle is less than one, therefore the deadtime is not a concern (provided that it is lower than the reciprocal of the repetition rate). Spurious pulses (dark counts) could be more frequent than the repetition rate, but since pixels of the Ra sensors are built up of multiple SPADs, the pixels stay sensitive even when one of their SPADs is inactive (due to a dark count). Also, this thesis targets time-resolved applications, in which a pulsed laser is used. As demonstrated in chapter 5, the Ra sensors cannot detect more than one photon per laser pulse on a pixel (even in SPC mode). Consequently, the deadtime of SPADs does not limit the achievable count rate, provided that it is lower than the laser period. No limitation was observed with the applied laser repetition rates (20 MHz in most cases, but up to 50 MHz for the Raman experiments in chapter 4), suggesting a pixel deadtime shorter than 20 ns. The PDE of Ra-I was not measured due to the complexity of accurate PDE measurements. It is expected, however, that it would match the values previously reported with the same/similar SPAD structures (Richardson *et al.*, 2011; Webster *et al.*, 2012b). Based on these SPAD structures, afterpulsing of the SPADs of Ra-I is expected to be negligible (Webster *et al.*, 2012b). Crosstalk was not measured quantitatively either, although tests were carried out providing qualitative assessment

of this phenomenon when using the red SPADs of Ra-I (which were designed for high efficiency at red and NIR wavelengths).

With the Ra-II sensor, characterisation of DCR, IRF, TDC resolution and masked/gated operation were carried out in this thesis. Some of these measurements were performed using modified versions of already existing program code and routines. For these, the author is thankful, especially for Dr. Ahmet Erdogan. Additional results of PDE and delay resolution measurements are also presented below. These measurements were carried out by Dr. Ahmet Erdogan (Erdogan *et al.*, 2019), using a different sensor chip from the same batch, relying on a measurement system at ST Microelectronics, Edinburgh.

3.3.1 Dark count rate

Ra-I

To measure DCR on Ra-I, the sensor was switched to SPC mode but no additional light source was used. Instead, the sensor and the PCB holding it were wrapped in several layers of blackout fabric (BK5, Thorlabs Ltd, UK) at room temperature, in a dark room with black walls and all electronic devices, except a PC, turned off. All light-emitting diodes (LEDs) on the PCB were either turned off or covered with blackout masking tape (T743, Thorlabs Ltd, UK). The exposure time was 1 ms, set through the sensor controlling firmware. 1000 lines of data (i.e. measured photon counts on each pixel of the line) were saved on the PC, to which the sensor was connected. The screen of the computer was turned off during the acquisition of the data. The 1000 lines of photon counts were summed on the computer (using Matlab 2017b, MathWorks Inc, USA), to represent the dark count rate per pixel in CPS unit. With the given experimental conditions, sources of dark count noise other than the sensor are considered to be negligible. Several DCR measurements were taken while a range of different bias voltages was applied to the SPADs through the firmware, starting from the breakdown voltage in steps of 250 ms. DCR measurements were carried out for both red and blue pixels.

Ra-II

The DCR of Ra-II was measured in a similar fashion as with Ra-I, except that DCR was measured for each SPAD instead of for each pixel. This was carried out by enabling one SPAD per pixel at a time, which is not possible with Ra-I. DCR measurements were carried out at different bias voltages starting from the breakdown voltage in steps of 500 mV. The exposure time was 500 ms in each case, and the acquired counts were divided by the exposure time on the host computer (using Matlab), to get DCR values expressed in CPS.

3.3.2 Instrument response function

Ra-I

For the IRF measurements (Fig. 3.8), a supercontinuum laser (WhiteLase micro, Fianium Ltd, UK) of 20 MHz repetition rate was used as a light source, filtered to a single wavelength of 2.5 nm bandwidth by a tuneable filter unit (Laser Line Tunable Filter, Fianium Ltd, UK). The average optical power varies with the selected wavelength, at 520 nm it was measured to be roughly 20 µW, using a power meter (PM100D with S130C, Thorlabs Ltd, UK). Free space collimated output of the filter unit was coupled to a fibre patch of 600 µm core diameter and 0.39 NA (numerical aperture) (M29L01, Thorlabs Ltd, UK) with a focusing lens of 50 mm focal length (AC254-050-A-ML, Thorlabs Ltd, UK). The light exiting the fibre patch was attenuated by 6 orders of magnitude with a neutral density (ND) filter (NE60A-B, Thorlabs Ltd, UK), in order to avoid early photon pile-up and distortions of the measurement. Measurements in SPC mode showed that the rate of detected photons from the attenuated laser light was less than 1 % of the laser repetition rate. The sensor was switched to TCSPC mode, and a total of 102 400 lines of TCSPC data (1024 lines in 100 iterations) were recorded for each wavelength, as the filter iterated through a range of wavelengths. Each line of TCSPC data, i.e. the arrival time of the first photon at each pixel, was acquired with an exposure time of 100 µs, set through the firmware. The acquired timestamps were built up to histograms on the host computer, for each pixel and wavelength individually. The jitter time profiles given by the histograms were further analysed using Matlab. For this, the timepoints of the time profiles were determined first, using the measured TDC resolution of each pixel (see below). After that, half of the peak intensity was determined on each time profile. In most cases the exact half-maximum intensity values are not detected in any timepoint (due to the quantised nature of the time axis). Therefore, the estimated timepoints, where half of the maximum intensity would be detected, were determined using linear interpolation between the intensity values of the two closest timepoints. The estimated timepoints with half-maximum intensity were used for full width at half maximum (FWHM) characterisation of the jitter time profiles.

The temporal shape of the incident laser pulses was not measured. In fact, the described method cannot separate the timing uncertainty of the sensor and the temporal shape of the laser pulse. Since the expected pulse length is in the picosecond range, while the sensor timing jitter is a few hundred of picoseconds, the effect of the finite laser pulse width on the IRF (and any potential jitter caused by the ND filter) is not significant. Furthermore, the measured IRF values can be used to describe timing uncertainties in any application, where the same light source and detector are used, and no additional timing jitter is expected from the optical components.

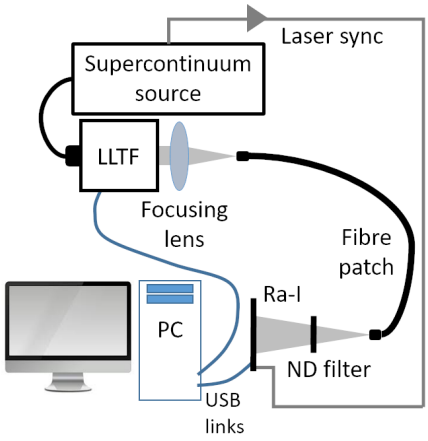

Figure 3.8: Arrangement for measuring the instrument response function of Ra-I. LLTF - Laser Line Tunable Filter, ND - neutral density

Ra-II

In preparation of using the Ra-II sensor for a specific application, the IRF was measured in a bigger optical setup including a fibre coupler based interferometer and a spectrometer, deploying Ra-II. Details of this optical system are specified in section 5.4.1. Filtered light of 38 nm FWHM around 850 nm centre wavelength from the Fianium supercontinuum laser was dispersed on the 512 pixels of the sensor. The power entering the interferometer was 950 µW, measured with a power meter (PM100D with S130C, Thorlabs Ltd, UK). The light in the spectrometer was attenuated by an ND filter (NE20A-B, Thorlabs Ltd, UK). In the case of the blue SPADs, histograms of photon arrival times were generated with the on-chip histogram mode of the sensor, using 32 histogram bins. The histogram bin width was measured in a separate experiment (TDC resolution measurement, see below). The IRF peak was positioned over the time-range covered by the histogram bins, by delaying the laser sync signal (STOP signal of the TDCs), using an analogue delay box (DB64, Stanford Research Systems, Inc., USA). 1 line of histogram data was acquired with an exposure time of 1 ms.

The histogram mode with variable bin width but fixed number of bins offers a trade-off between the timing accuracy and the length of the analysed time-interval. In order to capture the long tail of the response of red SPADs with high timing resolution, red SPAD IRF was measured in raw TCSPC mode. External delay was applied to the STOP signal (laser sync) in order to shift the laser peak to low timestamps and avoid broadening of the IRF due to jitter of the TDCs. A higher attenuation of the incident light was required in this case (NE60A-B, Thorlabs Ltd, UK), in order to avoid early photon pile-up. The applied exposure time was 100 µs and 100 000 TCSPC lines were recorded. Timestamp histograms were built on the computer (using Matlab).

For both sets of data, the FWHM of the peak on the histograms was calculated. The estimated timepoints belonging to half of the peak maximum were determined by using interpolation between the closes timepoints, as discussed in a previous paragraph.

3.3.3 Time-to-digital converters

Ra-I

In the time-resolved modes of Ra-I, the timespan between a photon absorption pulse and the STOP pulse is measured by stepping the respective counter with a certain time period, set by the ring oscillators (see Fig. 3.4). In order to turn these digital timestamp codes to real time information, the time period at which the counter is stepped has to be known. This period is also the shortest time that the TDC can measure, and therefore referred to as the TDC resolution. The resolution of the TDCs can be estimated by analysing the differences of the digital timestamps in time-resolved measurements of events with a known time difference.

For measuring the TDC resolution, a pulsed laser (44 ps pulse width, 403 nm laser head, PLP10 laser driver, Hamamatsu Photonics K. K., Japan) was used as a light source (Fig. 3.9). The lens of the laser head was intentionally misaligned, so that the diverging beam covered the entire sensor. The sync signal of the laser was connected to the sensor through a digital delay generator (DDG) (DDG645, Stanford Research Systems, Inc., USA). The sensor was used in TCSPC mode to measure the time between absorption of photons of the laser light and the sync pulses. Using the TCSPC mode required that the intensity of the laser is set to a low value on the driver (providing an average power of 120 nW for the whole beam, of which only a fraction was falling to the sensor), so that the detected photon count rate was lower than 1 % of the laser repetition rate. This was confirmed with measurements in SPC mode. For each pixel, histograms of photon arrival times were generated in a similar way as in the IRF measurements, from 1500 lines of TCSPC data, each taken with an exposure time of 100 µs. At each pixel, the time bin with the peak of the histogram (i.e. the most frequently occurred timestamp) was used as the measured value of the time between photon absorptions and the respective laser sync pulses, given by the pixel's TDC. The real time between these two is not known, and therefore cannot be used to map the digital timestamp code to real time. Instead, an additional delay of 15 µs was added to the sync pulses, using the DDG. The added delay increased the time between photon absorptions and the respective sync pulses, with a known amount. Another set of histograms were generated from measurements with the same parameters (regarding the exposure time and number of TCSPC lines), but with the additional delay. The bin with the highest intensity (most frequent timestamp) was noted again, for each pixel. TDC resolution of a pixel was defined as the known delay of 15 µs divided by the

difference of peak bins (timestamp difference) in the two, delayed and non-delayed cases.

The described method for TDC resolution measurement limits the attainable accuracy, due to the finite number of timestamp bins between the peak bins of the delayed and non-delayed histograms. To put it in another way, the added delay is not always integer multiple of the TDC resolution, whereas the time difference of the most frequent timestamps inevitably is. In the worst case, the added time delay is almost as long as the difference of $N+1$ timestamps would be, but only causes a timestamp difference of N on the histograms. The above defined TDC resolution in this case will be $\frac{N+1}{N}$, which has an error of $\frac{1}{N}$ with respect to the true TDC resolution. The 15 µs delay was chosen to be long enough to span over several thousands of time-bins, in order to minimise this quantisation effect. Accordingly, the repetition rate of the laser was set to be a low value of 50 kHz, in order to avoid 'wrap around' of the delayed peak on the histograms. At this repetition rate, and even lower detected count rate (required by TCSPC), the red SPADs had to be used with much lower DCR, so that the measured signal did not fade in noise. It should be noted as well, that any constant delay caused by optical components or the detector is cancelled out in the measurement, owing to the differential nature of the measurement. Timing uncertainties and variations in the recorded timestamps are not a concern either, provided that a sufficient number of arrival times are considered for the histogram, so that the most frequent timestamp can be selected with high certainty.

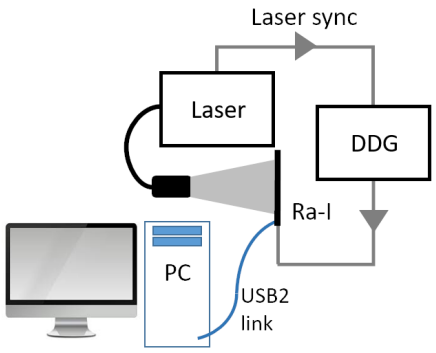

Figure 3.9: Arrangement for measuring the time-to-digital converter resolution of Ra-I. DDG - digital delay generator

Ra-II

The TDC resolution of Ra-II was measured in raw TCSPC mode of the sensor. Similarly to the IRF measurements, filtered light of the supercontinuum laser (38 nm bandwidth around 850 nm centre wavelength, and 950 µW power) was shone onto the sensor. The light was attenuated to provide an adequate photon count rate for TCSPC operation (avoiding pile-up), using an ND filter (NE60A-B, Thorlabs Ltd, UK). Measurements in SPC mode confirmed, that the incident count rate was less than 5 % of the laser repetition rate. The 20 MHz sync output of the laser was connected to the sensor through the DB64 analogue delay box, as STOP signal for the TDCs. TCSPC histograms were built from several lines, acquired in two cases. In the first case, the external delay was set so that the peak of the laser pulse on the histograms appeared at low TDC values. In the next case, an additional delay of 43 ns was applied to the STOP signal, shifting the histogram peak to high TDC codes. In each case, 50 000 TCSPC lines were acquired with 300 µs exposure time for each. Histograms were built from the recorded timestamps for each pixel, on the host computer. The known delay between the two cases was divided by the difference between the bins with highest population in the histograms to estimate the resolution of each TDC.

3.3.4 Photon detection efficiency

Ra-II

PDE measurements were carried by Dr. Ahmet Erdogan on a separate Ra-II chip (Erdogan *et al.*, 2019). The wavelength was changed from 350 to 1050 nm in 10 nm steps. To rule out the effect of crosstalk in the results, one SPAD was enabled per pixel at a given time. The ratio of the number of detected and incident photons was calculated for each SPAD. While the first one is simply the output of the sensor, the latter one can be calculated using the known optical power falling onto the device. PDE was calculated as the average of these ratios over the SPADs.

3.3.5 Time-mask, time-gate and delay resolution

Ra-I

Time-masking feature of Ra-I can be used to restrict counting of photons to a narrow temporal region. This is particularly useful when the signal of interest arrives at a certain time interval (with respect to the light pulses of the source), while high background noise, which is uncorrelated with the light source hinders the detection of the weak signal. In theory, this is also possible by using a short exposure time which is started right at the arrival of the signal of interest. However, such short exposures repeatedly started at specific timepoints is challenging. With a time-gate or time-mask,

set at the right timepoint (with respect to the source pulses), the narrow time region can be measured multiple times throughout a long exposure. The amount of noise excluded depends on the width of the applied time-mask. Therefore, the resolution of the position and width of the time-mask is an important parameter when evaluating the efficiency of a masking scenario. To that end, the masking performance was tested both in the TCSPC and SPC modes of Ra-I. In light of the discussion about the how the timepoint of the mask edges are defined (see Fig. 3.2), the masking performance can be linked to the resolution of the delay line, and the measurement of the delay of a single delay element.

Time-correlated single-photon counting The aim of these measurements is to estimate the resolution of the delay line and the shortest applicable time-mask, using TCSPC mode of the sensor. First, flat TCSPC histograms are generated from equally frequent photon arrival times. Applying various delays to the rising falling edges of the time-mask (Fig. 3.10a) restricts the recording of photon arrivals to a limited time region, and results in windowing the flat time profiles (Fig. 3.10b). The mask resolution is to be estimated by measuring the width of the windows.

In these measurements the sync output of the PLP10 laser driver was connected to the sensor to provide STOP signal for the TDCs and sync signal for time-mask generation, however no laser head was attached to the driver. To maintain a low level of light uncorrelated to the sync pulses, a filament desk lamp was placed far from the sensor. Measurements in SPC mode were carried out to confirm that the rate of photon detection was less than 1 % of the 1 MHz repetition rate of the sync pulses, to avoid the over=representation of early photons (pile-up, see section 2.1.3) and counting loss. The position of the rising and falling edges of the time-mask (with respect to the sync signal) were generated using the on-chip delay line of the sensor (Fig. 3.10a). Histograms of photon arrivals were generated from 200 000 TCSPC lines, each recorded with an exposure time of 50 µs. Several histograms were collected, while different positions of the time-mask were set by selecting the number of delay steps for the mask edges. 'On' windows (where photon processing is only enabled for a short period) were generated using 10 delay steps for the rising edge and 11 delay steps for the falling edge, then 20 delay steps for the rising edge and 21 delay steps for the falling edge, and so on. These settings of the mask edges result in time-masks with a width of 1 delay step, but at different positions along time. After that, several 'off' windows were used as well (in which photon processing is enabled any time but within the window). 'Off' windows were set to have 1 delay step width as well, at different locations, i.e. 10 delay steps for falling edge and 11 delay steps for rising edge, and so on. The FWHM of the windows were evaluated as described before, using interpolation to find the approximate timepoints where the window height is half of its maximum.

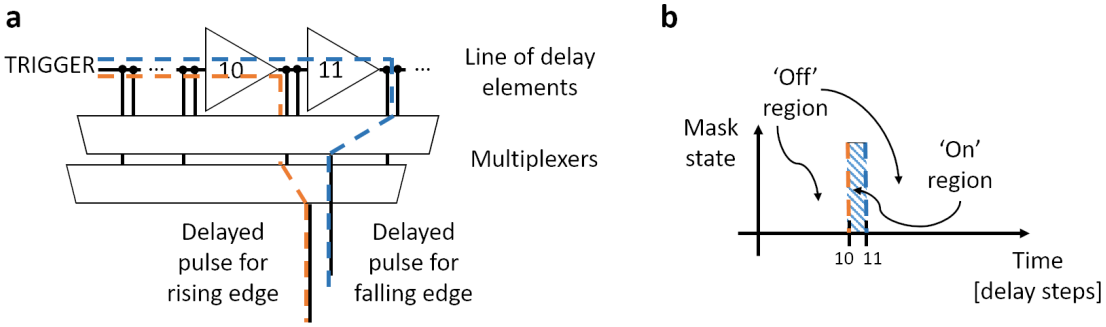

Figure 3.10: Defining the width and position of a time-mask, by delaying a trigger signal separately for the rising and falling edges of the mask (a). The trigger pulse is stepped through a range of delay elements on the delay line. Multiplexer units select the amount of delay by connecting one of the nodes between delay steps to their output. The delayed pulses travel to a dedicated unit, in which a time window is formed by blocking the SPAD pulses that are outside the mask period (b).

Single-photon counting In SPC mode (Fig. 3.11), a short time-mask window with a width of 1 delay step was set. The blue pixels of sensor were illuminated with pulsed laser light, using the PLP10 driver and the 483 nm laser head. The power of the optical signal was measured to be 70 µW, which was further attenuated by an ND filter (NE40A-A, Thorlabs Ltd, UK). The laser was externally triggered by a 5 MHz signal, generated by the DG645 delay generator. The delay generator also provided the STOP signal for TDCs by delaying the 5 MHz trigger signal at a separate channel. Changing the delay between the two channels results in changing the position of the time-mask relative to the laser pulse. 500 lines of 100 µs exposure time were recorded in SPC mode and the total photon count (sum of the 500 lines) was calculated at each delay between the two trigger signals. Scanning through a range of delays between the channels in 5 ps steps allowed to move the time-mask across the laser pulse.

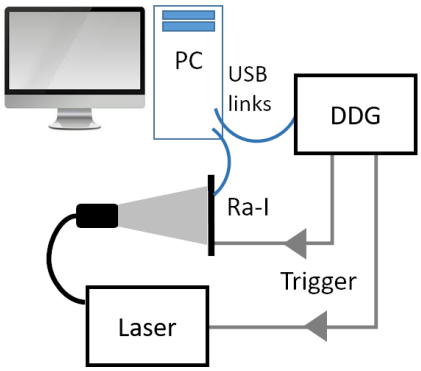

Figure 3.11: Measurement arrangement to reveal the time mask profile in single photon counting mode, by sweeping a laser pulse in the mask window with a delay generator. DDG - digital delay generator

Ra-II

Time-mask Time-masking feature of the Ra-II sensor was tested with both types of pixels, in the same optical setup as with the previous characterisation measurements of Ra-II. Various positions and widths of the two time-masks were tested, by delaying the rising end falling edges of the masks through the four delay channels of the sensor. The time profile of each time mask was measured by sweeping the mask position with respect to the timepoint of the laser pulses, as described before. The applied light source was the filtered Fianium supercontinuum (38 nm FWHM around 850 nm centre wavelength, and 950 µW power attenuated with an ND filter (NE20A-B, Thorlabs Ltd, UK). A 5 MHz trigger signal was used to initiate delay generation, which was provided to the sensor by pre-scaling the 20 MHz sync output of the Fianium supercontinuum laser, using the DDG645 delay generator. Pre-scaling of the signal was required by the DDG, as it can only handle repetition rates below 10 MHz. When the trigger signal is pre-scaled with respect to the optical signal, several optical pulses hit the sensor in the trigger period (Fig. 3.12). However, photons are only detected when an optical pulse reaches the sensor in the 'on' region of the time-mask. Assuming that the mask is not too wide to detect two or more laser pulses in the trigger period, sweeping the mask position (trigger signal) with respect to the timepoint of laser pulses results in the convolution of a laser pulse and the time-mask. The number of measured photons in function of the delay from the DDG gives the time-profile of the 'on' mask. 'Off' masks cannot be profiled this way, since several optical pulses are detected by the sensor in the trigger period, independently of the mask position. This expected behaviour of the 'off' time-masks were confirmed through measurements as well (not shown).

With each time-mask setting on the sensor, the DDG iterated through a number of delays in 5 ps steps, effectively sweeping the time-mask across the light pulses. At each position, 50 lines were acquired in SPC mode, with 100 µs exposure time. The 50 lines were summed on the computer, and for each pixel the photon counts were plotted in function of the applied delay of the mask position to reveal the mask profile. FWHM of the mask profiles were calculated, using interpolation, as previously described.

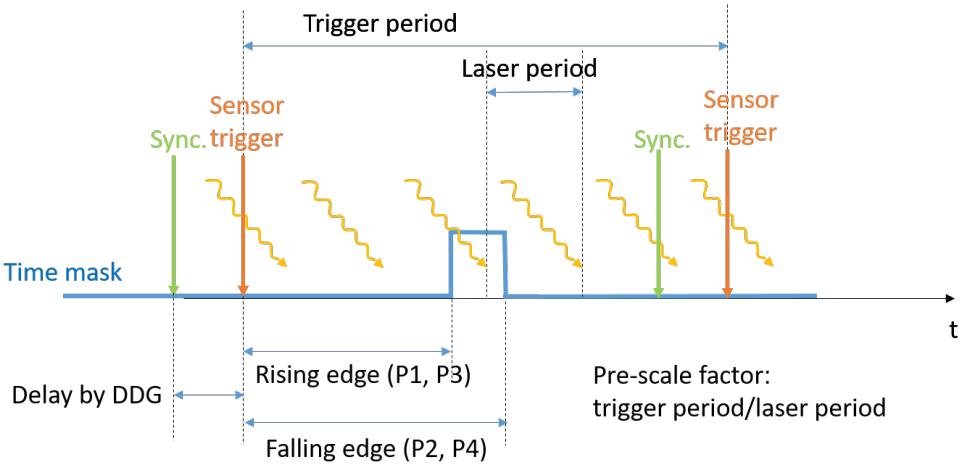

Figure 3.12: Relative positions of optical and electrical pulses when profiling 'on' time-masks and time-gates.

Time-gate The time-gating feature with blue SPADs was analysed in the same fashion as the 'on' time masks, but in this case the sensor was configured to use a time-gate instead of a time-mask. Various gate sizes and positions were tested by applying the appropriate delay to the gate edges.

Delay resolution

Characterisation of the delay generator on a separate Ra-II sensor was performed by Dr. Ahmet Erdogan. A sync signal, starting delay generation, and the output of a delay channel of the sensor were connected to a high resolution oscilloscope (WaveRunner 640Zi, Teledyne LeCroy, USA). Increasing delay codes were applied to the sensor, and the generated delay was determined by the oscilloscope, by measuring the time difference between pulses of the two channels. Details of the methods used for the measurement is given by Erdogan *et al.* (2019).

3.3.6 Crosstalk

Ra-I

To test if the pixels of Ra-I are prone to crosstalk, an optical system was used built by Ehrlich *et al.* (2017a). The optical arrangement consisted of a Fianium supercontinuum laser, filtered with an LLTF filter unit (Fig. 3.13). The selected wavelength from the filter unit was 500 nm, coupled in to a fibre patch, which was connected to a spectrograph. The spectrograph employed a spherical mirror for collimating the input beam, a diffraction grating and another spherical mirror to focus the beam to a Ra-I sensor. The spectrograph did not play a crucial role in the crosstalk experiments. The focused spot

was centred at pixel 127 of the line. The diameter of the spot was estimated to be around 100 µm, based on the number of pixels covered by the spot and the known pixel pitch. The bias voltage was set close to the breakdown voltage of the SPADs, and increased in steps of 20 mV. At each bias voltage, 20 lines were recorded with an exposure time of 100 µs in SPC mode, then the laser was shuttered and another 20 lines were recorded. Each set of 20 lines was summed on a computer. SPC lines from the measurements with shuttered laser were subtracted from the respective lines with the laser on, to account for background subtraction. Peaks on the resulting lines were characterised by calculating their FWHM. For this, the points where the half maximum would be reached were estimated by linear interpolation between intensity values of pixels close the half maximum. The experiment was repeated at different wavelengths. The measurements were carried out together with Katjana Ehrlich, using a different Ra-I sensor than the one used in all the other characterisation measurements.

3.4 Results

3.4.1 Dark count rate

Ra-I

DCR of both red and blue pixels is shown in Fig. 3.14, at different bias voltages. The pixels along the *x* axis are sorted to show the DCR of pixels in an increasing order. The pixel with mean DCR yields around 1000 dark counts per second at 20.46 V bias voltage in the case of the red pixels, and 74 400 dark counts per second at 14.92 V bias voltage in the case of the blue pixels. Both voltages correspond to a bias voltage that exceeds the breakdown voltage of SPADs with 1.2 V (excess bias voltage). Median DCR in function of the excess bias voltage is shown in Fig. 3.15. Measured values are marked with squares, and the dotted line shows an exponential fit to the measured median DCR values (note the logarithmic scale), suggesting an exponential increase of DCR with bias voltage.

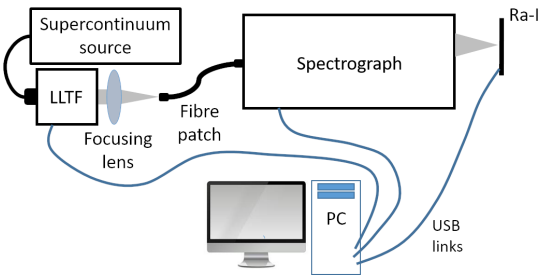

Figure 3.13: Optical setup used for characterising crosstalk of the Ra-I pixels.

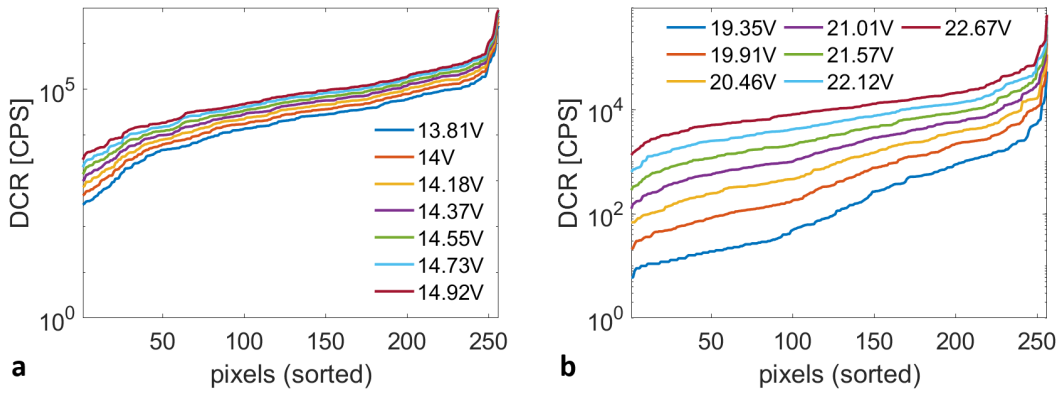

Figure 3.14: Dark count rate of blue (a) and red pixels (b) at different bias voltages (Kufcsák *et al.*, 2017).

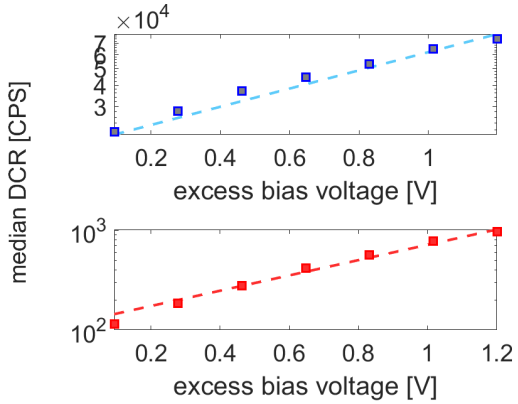

Figure 3.15: Median dark count rate versus excess bias voltage for blue (top) and red pixels (bottom) of Ra-I. Dashed lines are exponential fits to the data.

Ra-II

DCR of all blue SPADs at different bias voltages is shown in Fig. 3.16a. SPADs on the *x* axis are sorted according to their DCR. Median DCR of blue SPADs at 1.2 V excess bias voltage is 2230 CPS. On the Ra-II board, there are two types of red SPADs designed. Red SPADs of the bottom 4 rows utilize shallow trench isolation (STI) for isolating neighbouring SPADs. DCR of the top (without STI) and bottom red SPADs (with STI) is shown in Fig. 3.16b,c. SPADs without STI have a lower breakdown voltage, but at the same bias voltage level their DCRs is matching that of SPADs under STI, as seen in Fig. 3.17. Median DCR at 1.2 V excess bias voltage was measured to be 7.87 CPS for red SPADs without STI, and 8.07 CPS for red SPADs under STI.

Median DCR in function of the excess bias voltage is shown in Fig. 3.17. The dashed lines on the logarithmic plots show a good fit, matching the expectations of an exponentially increasing DCR with the bias voltage.

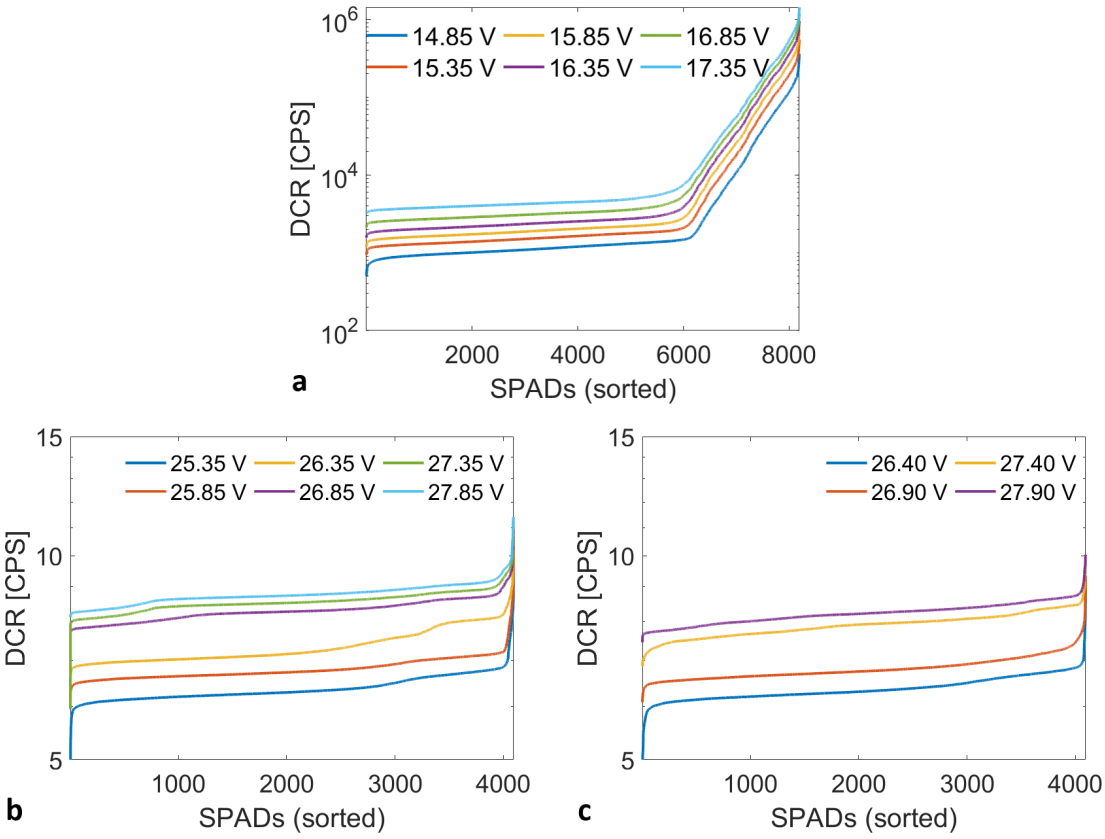

Figure 3.16: Dark count rate of the blue (a) and the red SPADs (without shallow trench isolation (STI) (b) and with STI (c)) of Ra-II at different bias voltages.

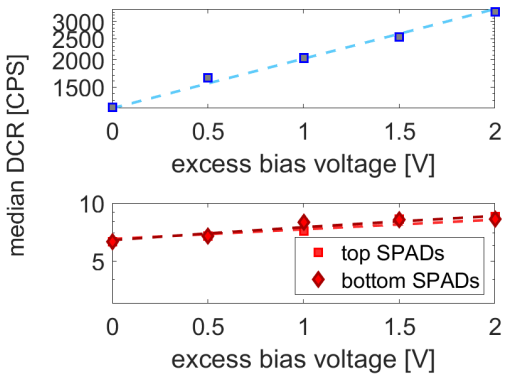

Figure 3.17: Median dark count rate versus excess bias voltage for blue (top) and red SPADs (bottom) of Ra-II. Dashed lines are exponential fits to the data.

3.4.2 Instrument response function

Ra-I

The IRF to laser light excitation of 600 nm wavelength is shown in Fig. 3.18 for both a red and a blue pixel (pixel 100). The red pixel yields an IRF width of 0.62 ns, whereas the IRF width of the blue pixel is 0.57 ns, both measured as FWHM. Wavelength dependency of the IRF is shown in Fig. 3.18 for blue (c) and red pixels (d), at 4 selected wavelengths (480 nm, 520 nm, 640 nm and 770 nm). The distribution of the IRF width across the pixels is shown using box plots.

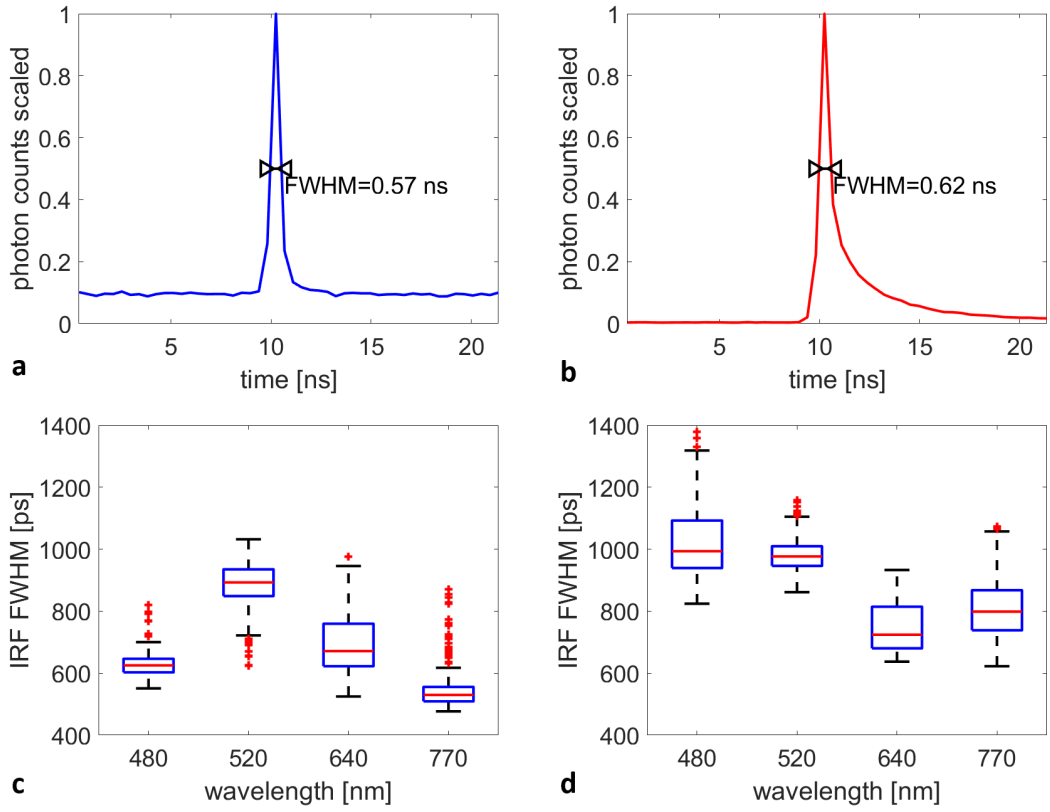

Figure 3.18: Instrument response function (IRF) to 600 nm wavelength laser excitation of a randomly chosen blue (a) and red pixel (b) (pixel 100 of the line). Blue pixels have a higher constant background due to higher dark count rate, but also a more accurate temporal response. The distribution of the IRF width across pixels is shown for blue (c) and red pixels (d), expressed as full width at half maximum. At each box, the central mark indicates the median value, the edges of the box indicate the 25 and 75 percentiles, the whiskers extend to the highest and lowest data points which are not outliers, and the '+' marks show outliers. Four different wavelengths of excitation were used: 480 nm, 520 nm, 640 nm and 770 nm. Revised from (Kufcsák *et al.*, 2017).

Ra-II

The IRF to 850 nm wavelength light for a blue and a red pixel is shown in Fig. 3.19. In order to directly compare the behaviour of the two types of pixels, the IRF of the blue pixel (acquired in histogram mode) was padded with zeros so that the two curves have the same *x* axis in Fig. 3.19. IRF of the red SPADs are heavily affected by a long tail.

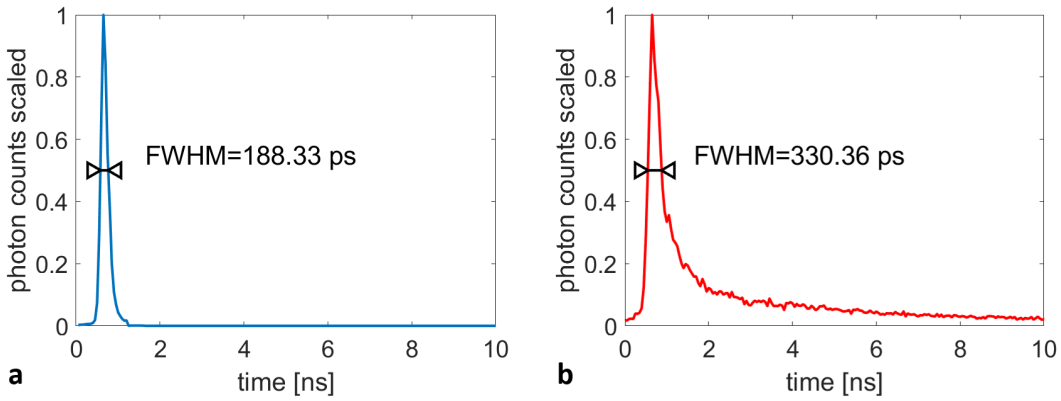

Figure 3.19: Sample instrument response function to 850 nm wavelength excitation through a complex optical setup, using a blue (a) and a red (b) pixel.

The IRF width of each pixel is depicted in Fig. 3.20 for red and blue pixels. The IRF with the blue pixels has a mean FWHM of 249 ps, with a standard deviation of 43.3 ps. When the red SPADs were used, three pixels/TDCs were failing to provide timestamps correlated to the STOP signal, resulting in excessively high FWHM of IRF values. Excluding these pixels, the mean FWHM is 303 ps, with a standard deviation of 42.1 ps.

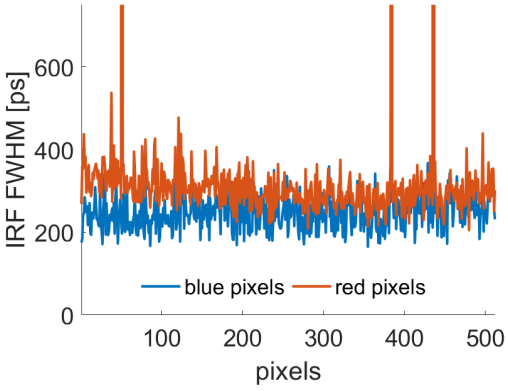

Figure 3.20: Full width at half maximum of the instrument response function with blue (a) and red (b) pixels of Ra-II. Laser light of 37 nm bandwidth centred at 850 nm was coming through an optical setup and dispersed onto the sensor.

3.4.3 Time-to-digital converters

Ra-I

The measured resolution of each pixel's TDC is shown in Fig. 3.21. The mean TDC resolution across all channels is 426 ps, with a standard deviation of 2.6 ps.

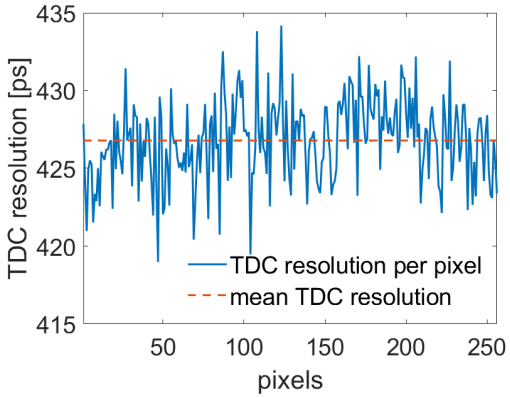

Figure 3.21: Resolution of time-to-digital converter (TDC) at each pixel of Ra-I. Mean TDC resolution is 426 ps (Kufcsák *et al.*, 2017).

Ra-II

Resolution of the 512 TDCs of Ra-II is depicted in Fig. 3.22. TDCs of three pixels did not function properly in the experiment. These pixels were left out from the plot. The mean resolution of the rest of TDCs was measured to be 50.4 ps, with a standard deviation of 0.5 ps.

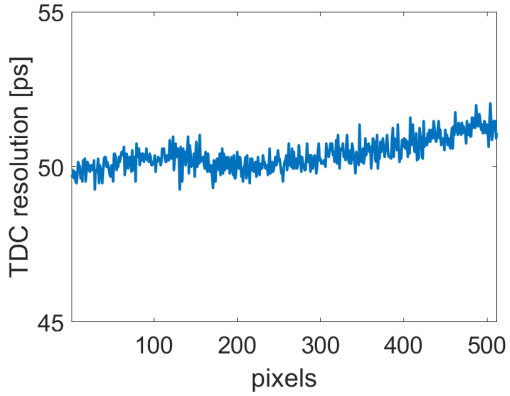

Figure 3.22: Resolution of the 512 time-to-digital converters on Ra-II. The mean time-to-digital converter resolution is 50.4 ns.

3.4.4 Photon detection efficiency

Ra-II

PDE for blue SPADs peaks at 480 nm with a value of 17 %, while the peak of the PDE for red SPADs is 7 % at 620 nm (Erdogan *et al.*, 2019), both measured at 1 V excess bias voltage (Fig. 3.23).

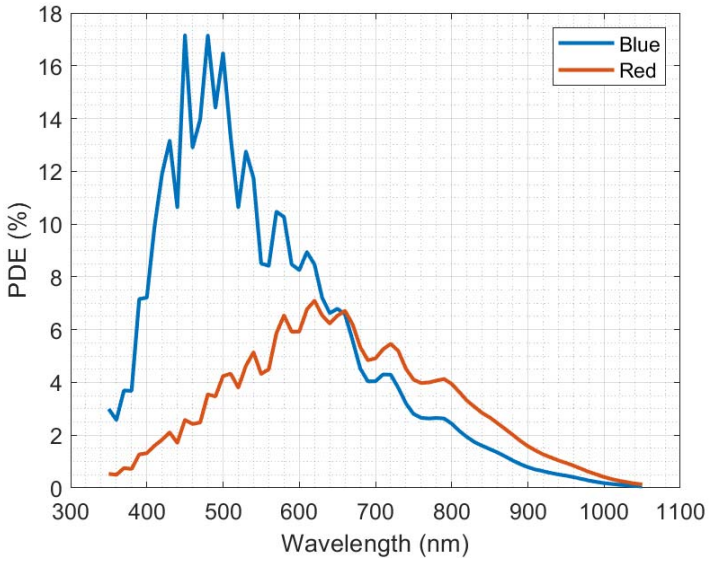

Figure 3.23: Photon detection efficiency of blue and red SPADs in function of wavelength, at 1 V excess bias voltage. Figure from Erdogan *et al.* (2019)

3.4.5 Time-mask, time-gate and delay resolution

Ra-I

Time-correlated single-photon counting The effect of 'on' (a) and 'off' masks (b) on TCSPC histograms at a representative pixel (pixel 100) is shown in Fig. 3.24. The width of the masked regions was measured when 10 delay steps were applied to the rising edge and 11 delay steps were applied to the falling edge of the time-mask. The average FWHM of the time window across pixels measures to be 1.44 ns in the case of the 'on' mask, and 1.56 ns for the 'off' mask.

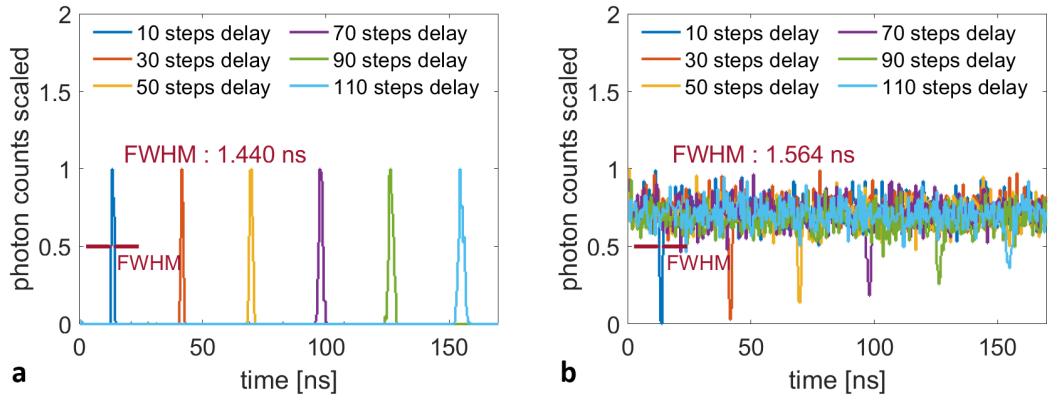

Figure 3.24: Time-masking time-correlated single-photon counting histograms of ambient light reveals the shape and size of the time window, and allows the measurement of the time mask resolution. Different positions of the mask is set by delaying its edges using the sensor's on-chip delay line. For the 10 step delay case, the average full width at half maximum across pixels was measured to be 1.44 ns for the 'on' mask region (a), and 1.56 ns for the 'off' mask region (b) (Kufcsák *et al.*, 2017).

Single-photon counting The number of detected photons in function of the delay between the time-mask and the laser pulse is depicted in Fig. 3.25. FWHM of the peak was measured to be 1.95 ns.

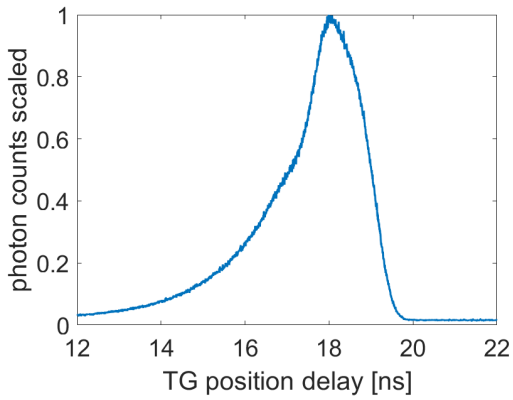

Figure 3.25: Sweeping the time point of the laser pulse and the position of the time-mask with respect to each other allows to reveal the shape of the mask window. The benefit of this method is that it provides much higher resolution of the window shape (compared to the measurement in TCSPC), however it shows a convolution of the time window and the laser pulse time profile. Here, an 'on' mask of 1 delay step width was swept with respect to the timepoint of the laser peak, using an external delay generator (Kufcsák *et al.*, 2017).

Ra-II

Time-mask Testing the time-masking features of the sensor revealed limitations when performing masked photon counting. Profiling was carried out with several time-windows of different positions and widths, for both of the two time-masks. The results showed that no stable masking can be performed when the difference between the rising and falling edges of the window is less then 13–15 delay codes (roughly 600–800 ps). Another limitation is related to the observation that the time masks tend to be more stable (with respect to width and counts outside the window) when the delay codes of the window edges are above 20–30. While, in general, a higher delay code infers higher accumulated jitter of the delay, due to the high resolution of delay step size (~50–60 ps) this is no major limitation.

Typical mask profiles with narrow widths (15 delay codes) are shown in Fig. 3.26 using a blue (a) and a red (b) pixel. The delay codes for the rising (P3) and falling (P4) edges were 23 and 38. Measured FWHM of the recorded windows suggests a time-mask resolution of 40–44 ps. The generated mean delay (and hence the window width) has some variation depending on the delay code, as indicated by Fig. 3.28(b).

Time-gate When it comes to time-gating, minimal time differences have to be maintained regarding the closeness of the edges of the SPAD enable signal (P3-P4 in Fig. 3.7), and the time between the rising edge of the SPAD enable (P3) and the falling edge of the quench signal (P1) for steady gating operation. Gate characterisation results show that the minimum delay code difference between turning the SPADs on (P3) and turning quenching off (P1) has to be around 280 codes, while the edges of the SPAD enabling period (P3-P4) has to be at least 290 codes apart. The minimal width of the gate window with such settings was measured to be around 7.8 ns on pixels at the edges of the sensor and 4.8 ns in the middle, both measured as FWHM. More uniform gate window across the pixels is achieved with slightly longer delaying of the gate edges. Figure 3.27 shows the gate window profile of a single pixel, with 20 and 450 delay codes for the rising and falling edges of the SPAD enabling signal, 400 and 450 delay codes for the falling and rising edges of the quench signal.

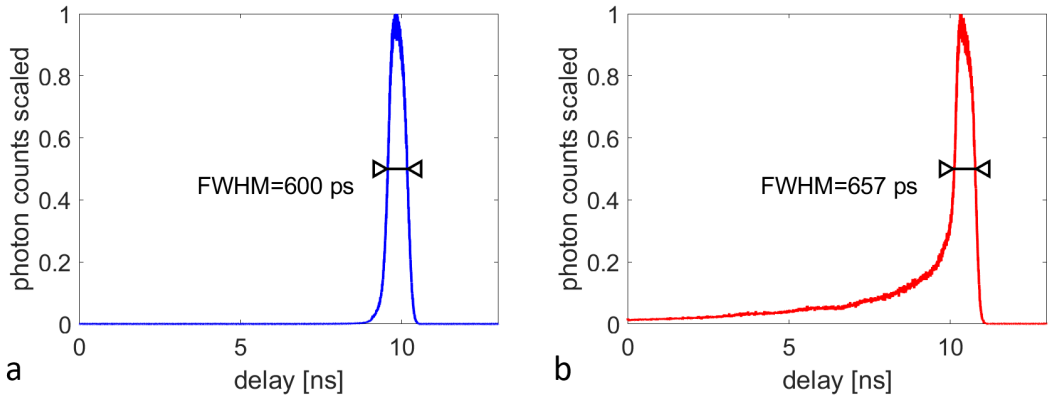

Figure 3.26: Time-mask profiles of Ra-II using blue (a) and red (b) SPADs (randomly selected pixel). The rising and falling edges were delayed with 23 and 38 steps with respect to the laser sync pulse, suggesting a delay resolution of 40 ps when measured with the blue SPADs, and 43.8 ns when measured with the red SPADs. The masks were profiled by sweeping a delay between the laser pulses and the trigger signal generating the time-masks.

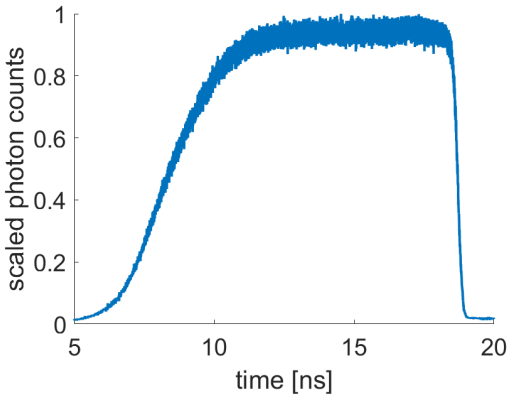

Figure 3.27: Time-gate window profile on pixel 256 of Ra-II, when the rising and falling edges of the SPAD enabling signal have a delay of 20 and 450 codes, respectively. Lowering the QUENCH signal was delayed with 400 codes, with respect to the laser sync signal, and the QUENCH signal was restored again with a delay of 450 codes.

Delay resolution Figure 3.28a shows the mean of the measured delay between delayed pulses and their trigger pulses, when a certain delay code is used. The error of the generated delay depicted in Fig. 3.28b is defined as the difference between the mean of the measured delays, and a linear fit to the measured values, describing the linearity of delay generation. The resolution of the delay generator (defined as the delay for 1 least significant bit (LSB) of the delay code) is measured to be 62.81 ps, based on the slope of the linear fit (Erdogan *et al.*, 2019).

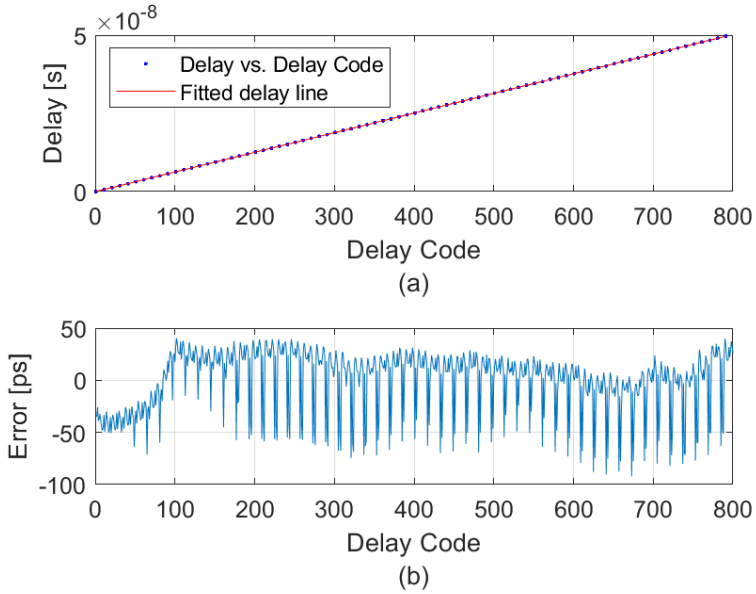

Figure 3.28: Mean measured delay versus delay code and linear fit to the measurements (a). Error is defined as the difference between the mean delay values and the linear fit (b). Figure from Erdogan *et al.* (2019).

Crosstalk

Widening of the laser peak due to more and more neighbouring SPADs breaking down with increasing bias voltages is evident when using the red SPADs. Fig. 3.29 shows the background subtracted SPC lines of the laser spot of 500 nm wavelength, at different bias voltages. In Fig. 3.30, the increase of the laser spot size is calculated as the FWHM of the peak, scaled to the spot size at the lowest bias voltage.

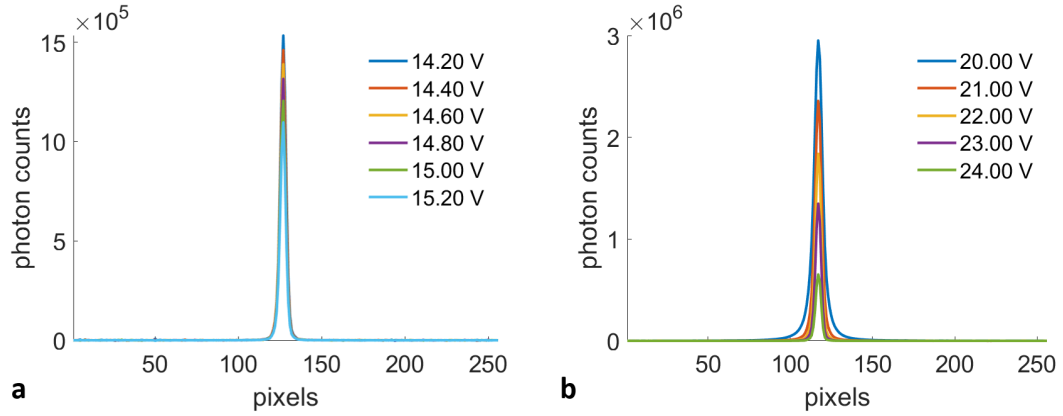

Figure 3.29: Photon counts acquired using the blue (a) and red pixels (b) at several bias voltages, when a focused laser beam of 500 nm wavelength is centred at pixel 127 of the Ra-I line sensor.

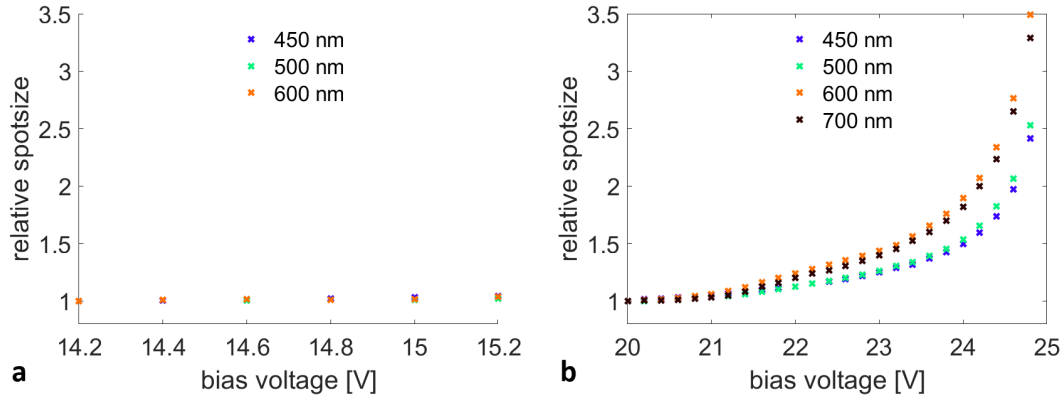

Figure 3.30: Increase of the detected laser spot size in function of the bias voltage for blue (a) and red pixels (b), for several different wavelengths of excitation.

3.5 Discussion

Ra-II was aimed to provide an improved performance with respect to Ra-I. However, its design was not driven by the characterisation results of Ra-I, since these measurements were only taken when the design of Ra-II was already finished. Still, performance of Ra-II indeed proved to be superior in almost all aspects. Measured characteristics of the Ra sensors are summarised in table 3.2. The first and obvious difference between the two sensors is the number of SPAD columns along the line (Ra-I - 256, Ra-II - 1024). More pixels mean a higher spatial resolution, which means a higher spectral resolution in spectral applications (and higher depth seen by the sensor in optical coherence tomography (OCT), see Chapter 5). More pixels, however, means a higher data size belonging to a line. This does not necessarily mean a lower frame rate, in fact, the higher number of output pads on Ra-II allows a higher data rate.

The price for this is a more complex physical connection between sensor and external circuits (on the PCB), and a more complex data pipeline from sensor to computer.

	Ra-I	Ra-II
Number of SPAD columns	256	1024
Number of SPADs per column	4	8
Median DCR @1.2V_{eb} [CPS]	74 400 (blue) 1000 (red)	2230 (blue) 8.07 (red)
Mean FWHM of IRF [ps]	627 (blue, @480 nm) 747 (red, @640 nm)	249 (blue, @850 nm) 303 (red, @850nm)
Number of TDCs	256	512
Mean TDC resolution [ps]	426	50.4
Peak PDE [%]		17 (blue, @480 nm) 7 (red, @620 nm)
On-chip data compression	CMM	Histogram mode

Table 3.2: Summary of the measured characteristics of the Ra sensors.

The number of SPADs per pixel is also higher on Ra-II (8 SPADs per pixel with 1024 pixels, or 16 SPADs per pixel with 512 pixels) compared to Ra-I (4 SPADs per pixel). The multiplexed pixel structure (i.e. several SPADs in OR connection) could increase the sensitivity of the pixel (turning an optical input to higher number of detected counts), and reduced pixel deadtime. On the other hand, it increases the DCR noise of the pixel. For a direct DCR comparison of the Ra-I and Ra-II sensors, we have to take into account that with Ra-I the DCR of each pixel with 4 SPADs was measured, whereas with Ra-II we have a DCR value per SPAD. Furthermore, the SPAD sizes differ on the two sensors. We may assume, that the probability for two SPADs of a pixel to fire at the same time is negligible, and approximate the median DCR of Ra-I SPADs as being one quarter of the median DCR of its pixels. Blue SPADs on Ra-II have a pitch of 11.890 µm, and a FF of 49.31 %. Ra-I blue SPADs have a pitch of 23.87 µm with 50.5 % FF. While the area of blue SPADs on Ra-I is 4.1 times bigger than on Ra-II, their estimated median DCR is 8.34 times higher than the median DCR of Ra-II SPADs. When comparing the DCR distribution of the two sensors across SPADs/pixels (Fig. 3.14a and Fig. 3.16a), we could further notice that Ra-II has much more uniform DCR distribution. However, the percentage of high DCR elements with Ra-II is somewhat higher than in the case of Ra-I. Ra-II red SPADs have a pitch of 13.69 µm and 14.54 % FF. The ratio of median red SPAD DCRs of the two sensors is 9.27, while the area of SPADs has a ratio of 31.25. Red SPADs of Ra-II are also more uniform regarding their DCR, compared to Ra-I. The distribution of DCR and, therefore, the shape of the curves in Fig. 3.14 and 3.16 depend on the applied manufacturing technology. The exact reasons, however, often not known or available for the SPAD

designers relying on a certain manufacturing technology and/or fabrication plant. The fact that Ra-II SPADs have lower DCR (even lower than suggested by their size), and that they are more uniform show a better noise performance of Ra-II, with both type of SPADs. The most probable reasons for this are the optimised design and manufacturing process. DCR noise can be even further reduced on Ra-II by turning off high DCR SPADs individually, although this may introduce a different type of noise in spectral measurements (Finlayson *et al.*, 2018).

The higher DCR of blue SPADs can be easily spotted in the IRF measurements of Ra-I as a higher background level, compared to the histogram taken with the red SPADs. The shape of the IRF is also quite distinct in the two cases. The IRF of red pixels has a much longer exponential tail, also known as 'diffusion tail'. This is due to the SPAD structure designed for enhanced detection of longer wavelength of photons. Photons absorption further away from the depleted region of the red SPADs generate charge carriers which diffuse randomly. These carriers can cause an avalanche if they reach the depleted region, but the time of the diffusion adds delay to the detection of the absorbed photon. Interpreting the IRF widths of Ra-I at different wavelengths (Fig. 3.18) is quite challenging, as the IRF measurements do not follow normal distribution over the pixels. Generally, it can be stated, that red SPADs yield a higher IRF width (due to the diffusion tail). If also considering the detection efficiency of the two types of SPADs at different wavelengths, e.g. by relying on Fig. 3.23, and assuming that the PDE values of Ra-I are similar, it can be seen that the range of IRF FWHM values (including outliers) increase at wavelengths where the efficiency of SPADs is lower. Ra-II pixels display a much narrower IRF, even though the two cases should not be directly compared, since they were measured in an entirely different optical arrangement. In fact, comparing the IRF measurement results of Ra-II with published measurements by Erdogan *et al.* (2019), we see that the published IRF widths are noticeably lower, indicating a significant jitter contribution from the optical setup in our case. A long IRF limits the accuracy of depth in time-of-flight measurements. It should be pointed out, that even with perfectly sharp laser pulses and no jitter of the applied optics, IRF measurements performed in TCSPC mode (or histogram mode, with Ra-II) inherently include the jitter of the timing circuits, not just that of the SPADs. Nevertheless, the IRF variation across the line is much lower with Ra-II than with Ra-I, although, uniformity of pixels should not be judged based on the results, as different pixels received different wavelengths of light in this measurement. Three red pixels yielded excessively high IRF width on Ra-II (Fig. 3.20). These pixels have faulty TDCs, so there was no peak in their respective time profiles. The FWHM of these time profiles is consequently meaningless.

As discussed, the demonstrated technique to estimate the TDC resolution is subject

to quantisation effects and errors. Owing to the high TDC resolution of Ra-II, delaying a signal causes a high difference in timestamps, even if the delay is relatively short. Therefore the TDC resolution measurement of Ra-II did not require long delays and a low repetition rate of the source. TDC resolution of Ra-II is about 8 times higher than that of Ra-I. Indeed, on Ra-I only the ripple counters of the TDC circuits are used (incremented after a full loop of 8 phases of the ring oscillators). In contrast, in Ra-II the 3 fine TDC bits are derived from decoding the state of the ring oscillator with respect to its phase when stopped. The achievable depth resolution of time-of-flight measurements is directly related to the TDC resolution. Also, the measurement of short temporal phenomena (such as the fluorescence lifetime of indigenous fluorophores) is impossible, if the LSB of the timestamp (TDC resolution) is longer than that. On the other hand, when measuring long events (e.g. phosphorescence decays), the timestamp counters may overflow with a fast timing circuit. In these cases a TDC with low resolution is more beneficial. The variation of TDC resolution across the pixels is also superior on Ra-II, with a standard deviation of 0.5 ps compared to 2.6 ps with Ra-I.

PDE measurement results confirm the intended difference in detection efficiency of the two types of SPADs (red and blue). An important feature to notice is the presence of fringes on the top of the PDE curves. This is believed to be caused by the etaloning phenomenon, in which interference occurs when light is reflected from surfaces along the vertical device structure (e.g. metallic wires). The interference depends on the distances between reflecting surfaces and the wavelength of the reflected light. Wavelength and distance dependent constructive and destructive interference appears as fringes along the spectrum. Etaloning affects the spectral content of the measurements, and has to be taken into account when accurate information is required on the measured spectral light intensities. PDE of the two devices cannot be compared, since no PDE measurements were carried out with Ra-I. However, PDE of Ra-I is expected to be similar to what was published before with similar structures (Richardson *et al.*, 2011; Webster *et al.*, 2012b).

The time-mask resolution of Ra-I was measured with two different techniques. Measuring the time profile of the window using ambient light and masked TCSPC has the benefit of revealing the true shape of the window. This is because time differences between the absorption of photons and their detection time is irrelevant, the time-mask simply cuts the flat TCSPC histogram according to its 'on' region. The downside of the technique (especially with Ra-I) is that its accuracy (sampling of the time profile) is set by the TDC resolution. Measurements in SPC mode with a shifted laser pulse can have a much higher accuracy of the profile (with respect to the timepoints in which it is measured), but what it really measures is a shape that is affected both

by the true window profile and the profile of the laser pulses. The reason for this is the following. In some cases, when charge carriers have a long diffusion time, triggering of the avalanche happens somewhat later than the absorbance of a photon. As a result, photon counts could be registered inside the specified time-window, even for optical pulses outside the time-window, resulting in a shape of time-windows as seen in Fig. 3.25. Another approach to explain the same phenomenon is to consider that the method follows the definition of a convolution (cross-correlation, more specifically), not showing the true time-profile of the time-window. This is the reason why the two methods resulted in slightly different values when measuring the time-mask window width of Ra-I. The measurements in TCSPC mode revealed an important effect, when long delays are used for the mask edges. As the mask edge delay is longer and longer, the accumulated jitter of the delay steps causes a broadening of the 'on' mask (Fig. 3.24a). Similarly, the accuracy of the 'off' mask edges reduces with increasing delays, hence the region where photons are not detected narrows (Fig. 3.24b). This could be avoided by using short on-chip delays, and shifting the mask to the desired position by applying a stable external delay on the TDC STOP signal.

Ra-II has a minimal window width that is half of that with Ra-I, and the step resolution of mask positions is 28 times higher. This is understandable, given the different delay structure (ring oscillator based). The reason why even shorter masks are not always stable is believed to be related with minimum time intervals, which have to be maintained for the propagation of signals within the timing structure. A higher resolution of the mask width allows narrower time-regions to limit the detection onto. This could lead to more efficient background rejection (e.g. in time-resolved Raman spectroscopy) or more flexible temporal restriction of unwanted signals (such as reflections in OCT, see Chapter 6). The window profile, when measured in SPC mode, is quite different for Ra-I and Ra-II. This could mean a different shape of the time-masks in the two cases, with Ra-II having a faster rise, and therefore more accurate edge. However, the applied light sources were different in the two cases, and this also affects the measured shape. It should be pointed out, that histogram mode of Ra-II would maintain a reasonably high sampling of the mask profile, while not being affected by any pulse and its shape (since using ambient light). The different shape of the mask profiles measured with the two types of SPADs on Ra-II is, again, a result of the diffusion tail of red SPADs. The shortest time-gate window of Ra-II is considerably longer than its shortest time-mask window. The benefit of gating is that photons are completely undetected in the off region of the time-gate, and hence strong bursts of light cannot saturate the sensor during this time.

The delay generator resolution measurements show a systematic deviation from a linear conversion from digital delay code to real time delay (Fig. 3.28b). Why the exact

reason for this is not known, it is suspected that it is related to the ring oscillator structure, in which the delay after a full loop is probably slightly different than the delay between phases within the loop. This should be no limitation, provided that the behaviour is consistent. However, no jitter measurement of the delay generator was carried out.

Finally, the crosstalk measurements clearly show a spread of SPAD activity across the neighbouring pixels with increasing bias voltage, when the red SPADs are used. This behaviour is a result of the deep structure of the red SPADs (Webster *et al.*, 2012b). The results do not reveal whether the source of crosstalk is optical or electrical, but nonetheless indicate that care should be taken when selecting the bias voltage of red SPADs.

Chapter 4

Time-resolved Spectroscopy with Ra-I

4.1 Introduction

Complementary metal-oxide-semiconductor (CMOS) single-photon avalanche diode (SPAD) based sensors are attractive in life sciences when low light levels are to be detected and time-resolved analysis is needed. SPAD sensors offer timing with tens of picoseconds resolution on several hundred parallel channels. They are also compact and robust against saturating optical signals (in contrast to photomultiplier tubes (PMTs), where high optical powers can easily damage the sensor). SPAD line arrays lend themselves naturally to spectroscopic applications, and have been demonstrated in various techniques, such as time-resolved fluorescence spectroscopy (TRFS) (Popleteeva *et al.*, 2015; Krstajić *et al.*, 2015; Ehrlich *et al.*, 2017a; Kufcsák *et al.*, 2017; Erdogan *et al.*, 2019), time-resolved Raman spectroscopy (Li and Deen, 2014; Ehrlich *et al.*, 2018; Nissinen *et al.*, 2018), cavity ring-down spectroscopy (Kiwanuka *et al.*, 2012), near-infrared optical tomography (Pavia *et al.*, 2015) and laser-induced breakdown spectroscopy (Maruyama *et al.*, 2014).

In this chapter, massively parallel time-resolved single-photon detection is demonstrated with the Ra-I sensor, which was the first available line sensor designed in our group. First, the instrument is described including the software, hardware and the sensor. The performance of the system is demonstrated in three applications: measuring the fluorescence kinetics of chlorophyll *a* in an intact leaf both in the time and spectral domains, measuring the fluorescence kinetics of a thrombin biosensor probe, and measuring the autofluorescence of excised murine lung tissue. As an indication of the system's applicability in time-resolved Raman spectroscopy, it is shown how time-resolved detection can be used to suppress noise when measuring Raman spectra. The above applications and the related results were published (Kufcsák *et al.*, 2017). The sensor system, including hardware and software, has been applied in further studies of Ehrlich *et al.* (2017a,b, 2018), which are described at the end of the chapter.

4.1.1 Time-resolved fluorescence spectroscopy with Ra-I

The measurement of fluorescence intensities is a crucial procedure of many applications in chemistry, biology and medicine (Valeur and Brochon, 2001). It is also known, that fluorescence lifetimes are characteristic for fluorophores and their environment (Lakowicz, 2006, chpater 1). The measurement of spectrally and temporally resolved fluorescence is beneficial for maximising information. Several instruments have been reported, where steady-state emission spectra and the time-resolved fluorescence signal of a single spectral channel can be acquired simultaneously (Ervin *et al.*, 2000; Pitts and Mycek, 2001; Pfefer *et al.*, 2003). The time-resolved fluorescence decay, in these cases, is detected by a sensor with fast response and short deadtime (typically microchannel plate (MCP) PMTs, or avalanche photodiodes). The time dependent electric output signal of the sensor is recorded using a digitiser or a digital oscilloscope of high sampling rate (Yankelevich *et al.*, 2014). Time-resolved data collection with pulse sampling at multiple wavelengths was achieved by means of a scanning monochromator and sequential decay measurements of several wavelengths (Marcu *et al.*, 1999; Maarek *et al.*, 2000; Fang *et al.*, 2004; Butte *et al.*, 2010). A different approach applied dichroic mirrors and fibre patches of different lengths in order to separate emission pulses of different spectral bands in time. Fluorescence emission decays of 3 (Sun *et al.*, 2008, 2011) or 4 different spectral channels (Xie *et al.*, 2012; Bec *et al.*, 2014; Yankelevich *et al.*, 2014) were detected using a gated MCP PMT, followed by sampling on a high-speed digital oscilloscope. This chapter focuses on building on the multichannel time-to-digital converter (TDC) structure of Ra-I, for massively parallel timing of fluorescence photons and the retrieval of spectral fluorescence lifetimes.

Efficient photon processing and sufficiently fast acquisition of fluorescence signals allows to observe how the fluorescence lifetimes change dynamically. Changes in fluorescent lifetimes due to resonance energy transfer and quenching can be used to study interactions with fluorescent probes, chemical reaction mechanisms and conformational changes in molecules, such as protein folding (Suhling *et al.*, 2015). Therefore, much attention has been paid on the construction of time-resolved fluorescence instrumentation to detect rapid changes in fluorescent lifetimes. Such systems perform 'double kinetic' experiments (Ishay *et al.*, 2012), where fluorescent decays (kinetics of probe fluorescence on picosecond and nanosecond timescale) are recorded at time points during the process of interest (kinetics of the chemical reaction on microsecond and millisecond timescale), as depicted in Fig. 4.1. Pulse sampling techniques are capable of rapid measurement of fluorescent decays, typically averaging several emission pulses for high signal-to-noise ratio (SNR). As a result, a high sampling rate of the observed reaction kinetic is possible, and a few number of sample reactions required to measure the transients, favouring sample preparation. However, this capability comes

at trade-offs. The response time of the detector and the sampling rate of the digitiser draw limitations on the observable lifetimes and the temporal resolution of the measured decay. The light source should provide short and strong pulses with energies in the order of 10 or 100 µJ. The high-end, high sampling rate digitising oscilloscopes and the amount of the associated data compromises multichannel detection (Ervin *et al.*, 2000). In the case of wavelength scanning techniques over many wavelengths, the limiting factor is the increasingly long acquisition time.

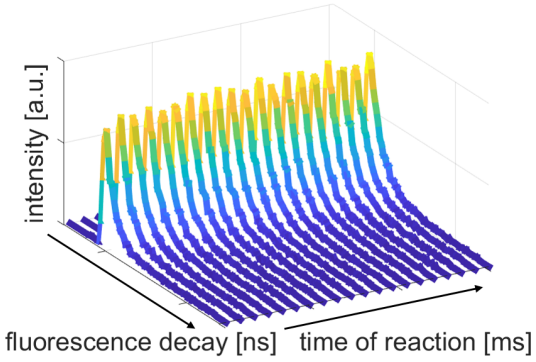

Figure 4.1: In double kinetics experiments the recorded data represents fluorescence along two timescales. Fluorescence decays with pico- or nanosecond resolution on one axis display the lifetime of fluorescence, while changes of the fluorescence lifetime and intensity are depicted on the second axis with micro- or millisecond resolution, as the chemical reaction progresses.

The benefit of the massively parallel timing capability of SPAD sensors, such as Ra-I, is that they detect the timing information at several hundreds of channels at the same time. The timing information in this case is not revealed by direct measurement of the decays (as in the pulse sampling technique), but through time-correlated single-photon counting (TCSPC). TCSPC is a widely used technique for fluorescent lifetime measurements in the time domain, providing exceptional temporal resolution (Lakowicz, 2006, chapter 4; Becker, 2015). Double kinetic measurements applying TCSPC were reported with millisecond sampling of protein folding reaction kinetics, using stopped-flow measurements (Beechem, 1997). To achieve sufficiently high SNR of the picosecond/nanosecond fluorescence decays, the reaction had to be initiated several hundred times. Histograms were built up from consistent time slices of the repeated reactions. Flow measurements with microfluidic mixers provided a sufficiently high number of photon detections, while maintaining a high sampling rate of the reaction kinetics (Kathuria and Bilsel, 2015). In this case, the sampling interval of the reaction time is governed by the spatial distances along the microfluidic channel. The number of collected photons and hence the quality of the decays are controlled by the dwell time over a single position.

TCSPC offers higher temporal resolution of the recorded fluorescence decays, compared to the pulse sampling method (Ervin *et al.*, 2000). However, TCSPC operation and the deadtime of the detector impose limitations with respect to the number of photons that can be processed. TCSPC also requires that the detected photon count rate is kept below a few percent of that belonging to the pulsed light source (Becker, 2005, section 7.9). Photon pile-up usually does not restrict photon-starved scenarios, but adds to uncertainty when bursts of fluorescence take place. In extreme situations, strong optical signals may need to be attenuated to avoid distortion of the measured temporal profile, thus increasing the measurement time. Multichannel TDC architectures are advantageous in this regard as well. Distributing the photon flux over several TCSPC channels can overcome this limitation by recognising multiphoton events on separate detection elements (Becker, 2005, chapter 3). TCSPC measurement of fluorescent lifetimes on the nanosecond timescale requires fast detector responses, for which PMTs have been used traditionally. The size of these detectors and the related timing instruments, however, prevents large array formats.

Massively parallelised CMOS SPAD arrays with several hundreds of pixels and per-pixel timing circuits offer highly efficient photon sorting capabilities with high timing precision. In recent years, several of such SPAD sensor arrays have been reported, targeting fluorescence lifetime imaging microscopy (FLIM) applications (Gersbach *et al.*, 2010; Veerappan *et al.*, 2011; Li *et al.*, 2011; Pancheri *et al.*, 2013; Krstajic *et al.*, 2015; Poland *et al.*, 2015), fluorescence spectroscopy (Krstajić *et al.*, 2015; Ehrlich *et al.*, 2017a; Kufcsák *et al.*, 2017), and spectral FLIM (Popleteeva *et al.*, 2015; Erdogan *et al.*, 2019). Our intention has been to show how time-resolved SPAD line arrays with several pixels and high line rate can advance applications in fluorescence spectroscopy, by measuring rapid changes and transients in spectral fluorescence intensities and lifetimes (Fig. 4.2). Measurement of fluorescence kinetics both in the wavelength and time domains has apparent advantages in the studies of Förster resonance energy transfer (FRET), reaction kinetics or protein folding dynamics, to name a few.

Wavelength scanning pulse sampling techniques are associated with rapid extraction of temporal information. The spectral information is retrieved over a longer measurement period. In contrast, the proposed method offers simultaneous spectral detection on hundreds of channels. Accordingly, it is the acquisition of temporal fluorescence information that is collected over a longer period of time. This approach offers greater flexibility in data acquisition. If fast sampling of the reaction kinetics is demanded, the detected counts in the spectral decays can be increased by binning several pixels along the spectrum or several time-bins in time. If high spectral resolution is required, one can use longer periods over which TCSPC data is recorded. The result is high

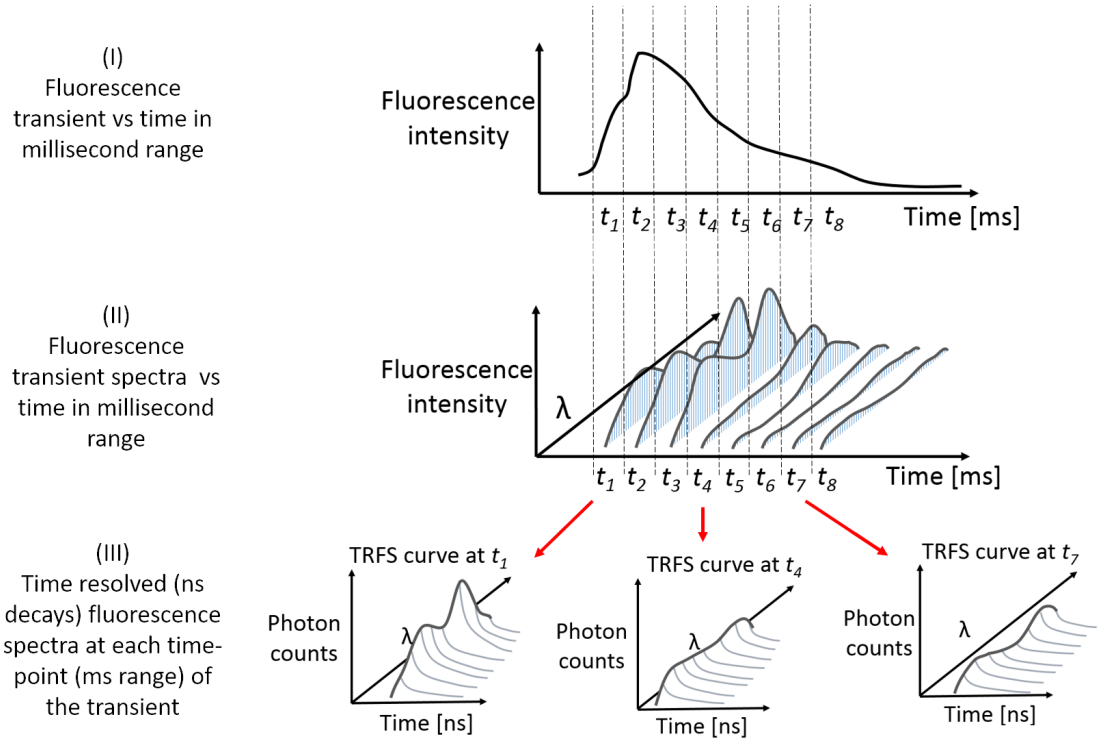

Figure 4.2: Measuring transients of spectral fluorescence intensity and lifetime. A fast photodiode can be used to detect changes in fluorescence intensity with millisecond and microsecond sampling (I). Compact spectrometers can be applied to measure changes in spectral intensity (II). Fast and efficient recording of photon arrivals on several pixels of our SPAD line sensor enables the acquisition of spectrally and temporally resolved fluorescence over a single transient, with millisecond sampling (III). The data from spectral time-correlated single-photon counting (TCSPC) is used to build histograms of spectral decays (TRFS datacubes (Kufcsák *et al.*, 2017) or wavelength-time matrices (Lloyd *et al.*, 2010)). The approach offers a trade-off between spectral resolution and photons per decay. Several pixels can be binned and longer TCSPC periods can be used for more counts per decay and increased accuracy of the lifetimes. These determine the spectral and temporal resolution of sampling, respectively. Figure from Kufcsák *et al.* (2017).

SNR decays at each pixel at the expense of the sampling interval used for tracking the observed reaction. The 'centre of mass' mode (CMM) provides even more efficient photon collection, and hence shorter acquisition times of temporal information, while maintaining high spectral resolution.

In many situations, repeated initiation of the reaction under study is not possible (e.g. studying *in vivo* enzyme kinetics). Our intention is to demonstrate a system, in which highly efficient photon time-stamping allows us to measure the spectrally and temporally resolved fluorescence data over a 'single shot' of the studied reaction. In addition to the high number of timing channels on Ra-I, this also requires a high

line rate of the sensor and comprehensive engineering of firmware and software. In the following sections, the Ra-I sensor system is described and its performance showcased in four applications. The aim with these example applications, on one hand, is to demonstrate the versatility of the proposed system, rather than focusing on the samples and using the results to analyse them. On the other hand, the results of the measurements are to be used to assess the system's capabilities, giving insight to the minimum sampling rates (both in time and wavelength) to still resolve the decays, and the extent to which trade-offs can favour timing or spectral aspects.

Fluorescence kinetics of chlorophyll *a*

Chlorophyll *a* exhibits a characteristic fluorescent transient kinetic (known as the Kautsky-effect) upon the induction of photosynthesis, as dark-adapted photosynthetic organisms are exposed to light (Maxwell and Johnson, 2000; Kalaji *et al.*, 2014). The effect typically displays a fast rise (within a second) in fluorescence intensity from zero to maximum level, followed by a slow relaxation (over several minutes). The Kautsky effect is connected to a variety of processes of photosynthesis and offers a non-invasive method to study many aspects of photosynthesis, the photosynthetic performance and ecophysiology of plants (Maxwell and Johnson, 2000; Stirbet and Govindjee, 2011). As such, the apparatus capable of spectrally and temporally resolving fluorescence kinetics over a single transient is of high importance in this field.

The study of chlorophyll fluorescence of leaves was chosen, since it is a typical case where repeating the reaction is problematic (the characteristic fluorescence is observed only after the leaves have adapted to darkness, which can take from 15 minutes to an hour (Stirbet and Govindjee, 2011)). Furthermore, upon excitation, fluorescence intensity shows a rapid increase over a second, requiring high sampling rate of the temporal evolution of the fluorescence signal. Commercial instruments provide tens of microseconds of resolution, but only measure the intensity of chlorophyll *a* fluorescence (Stirbet and Govindjee, 2011).

Fluorescence kinetics of a thrombin FRET probe

Förster or fluorescence resonance energy transfer (FRET) has been widely applied in biological disciplines to analyse molecular structures, cellular dynamics and live imaging of cellular processes, to name but a few (Lakowicz, 2006, chapter 13). Optical imaging probes are often based on FRET systems as well (Mills *et al.*, 2016). The application of fluorescing probes for *in vivo* diagnosis of lung diseases is a primary objective of the Proteus project (`https://proteus.ac.uk/`). The presence of bacteria inside the lung can be assessed through static detection of fluorescence from the associated probes. Additionally, measuring the dynamics of spectral fluorescence lifetimes offers

a tool for better understanding the underlying molecular structural changes, aiding the design of such probes.

For a practical demonstration of this, a FRET probe could be used, where the structural changes are studied through observation of dynamic fluorescence quenching. Changes in the fluorescence signals of these probes are expected to develop over several seconds. The low constraints on temporal sampling allows to record decays with high accuracy and over many spectral channels. Here, a custom probe for thrombin detection was used, readily available within the project, even though not producing complex fluorescence signatures.

Auto-fluorescence of excised murine lung tissue

Efficient photon collection and fast data acquisition is key for double kinetic experiments, as well as for confocal scanning microscopy. Frequent recordings of spectrally and temporally resolved fluorescence data could be done over spatial points (instead of timepoints). Therefore, a potential application for our time-resolved line sensor is confocal spectral FLIM, where spectrally and temporally-resolved data cubes can be recorded at each pixel of the FLIM frames. Such application is also related to the aims of the Proteus project, as it could provide *in vivo, in situ* medical diagnosis or a guiding platform for therapeutic interventional radiology. FLIM with reasonable frame rate is, however, challenging when using with TCSPC, especially if multiple wavelengths are to be detected (Yaseen *et al.*, 2013). The CMM offers rapid acquisition of temporal information. Assuming a measured photon arrival time for every laser pulse, and that the CMM value is accurate using 200 photons, the centre of mass of arrival times can be calculated in 10 µs at a single channel, with a laser repetition rate of 20 MHz. The task here is to investigate feasibility of video-rate spectral FLIM with Ra-I, and the necessary trade-offs between spectral resolution, size and rate of FLIM frames.

With such future applications in mind, the Ra-I sensor was tested to observe the fluorescence signal of tissue samples. In these experiments no spectral fluorescence lifetime transients are expected, still a similar acquisition protocol was followed as in the double kinetic experiments, to mimic the operation of a confocal scanning system (without the actual laser beam scanners). The sample was murine lung tissue prepared on microscope slides by Philip Emanuel.

4.1.2 Time-resolved Raman spectroscopy with Ra-I

Raman spectroscopy is a technique widely used for non-destructive analysis of the chemical composition and structure of materials. The technique is based on the detection of inelastically scattered light, shifted from the excitation wavelength according to the vibrational and rotational energy of the molecules in the sample (Ferraro, John R., Nakamato, Kazuo Brown, 2003). A major challenge in detecting the weak Raman signal is suppressing the several orders of magnitude stronger fluorescence, induced by the excitation light. Time-gating has been considered as one of the most powerful techniques to suppress the fluorescence background (Wei *et al.*, 2015). Time-gating in Raman spectroscopy utilises the fact that the Raman signal and the fluorescence background have different temporal behaviours. While the former occurs instantaneously with the laser peak, the latter one decays with its characteristic lifetime value (Fig. 4.3a). In addition, time-gating reduces the noise of the sensor in the recorded data. The information from time-gated Raman measurements of tissue can be used to complement fluorescence techniques. Such multi-modal system is of interest for medical diagnosis (e.g. in the Proteus project), for which time-resolved SPAD line sensors are convenient.

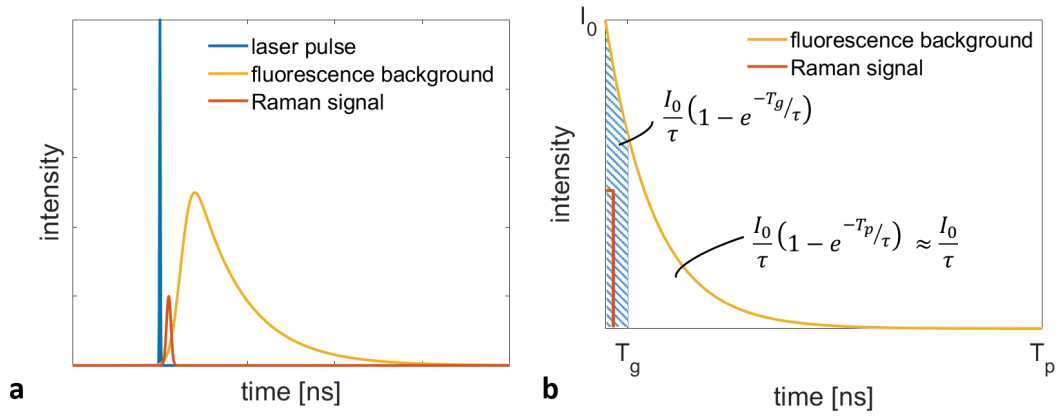

Figure 4.3: Illustration of the signal time profiles in Raman measurements (a). The amount of fluorescence reduction when time-gating on the Raman peak is estimated using the ratio of the total decay integral over the laser cycle period and the integral within the window. I_0 - initial fluorescence intensity, τ - lifetime, T_g - gate width, T_p laser period.

To estimate the effect of the time-gate width on noise and fluorescence suppression, we consider the Raman to noise and Raman to fluorescence ratios, with and without gating (equations 4.1). These ratios are defined with the integrals of the detected signals (R, F and N referring to Raman, fluorescence and noise, respectively, and the additional 'g' referring to the gated cases), which could be interpreted either by detected number of photons, or total energy measured in a laser cycle. The total detected

noise is the dark count rate (DCR) multiplied by the observation window width, that is the laser cycle T_p with no gating, and the gate width T_g with gating (equations 4.2). For the fluorescence signals, a single exponential decay model is used, with no timing jitter. The rise time of the fluorescence is also neglected (being in the femtosecond range), so that a narrow time window allows the detection of the Raman pulse, while rejecting the majority of the fluorescence decay (Figure 4.3b). The total fluorescence signals (without and with gating) are described with the respective integrals of the single exponential decay (equations 4.3, where I_0 is the initial fluorescence intensity, T_g is the gate width, and the approximation can be made if the laser cycle T_p is considerably longer than the lifetime τ). Since the noise is uniformly distributed in time, the Raman to noise ratio improvement is simply the ratio of the laser cycle width and the gate width (equation 4.4). The improvement on the Raman to fluorescence ratio equals to the ratio of the integrals of the total fluorescence decay and the time-gated decay (equation 4.5). From this, in the case of a fluorescence lifetime (τ) of 4 ns, a hundred fold improvement would require a gate width (T_g) of ~40 ps. Given the time-mask resolution of Ra-I, the improvement on the Raman to fluorescence ratio is expected to be moderate, also given the fact that the shortest applicable gate width is limited by the tail of the Raman peak, due to the TDC jitter.

$$r_N = \frac{R}{N}, \quad r_{Ng} = \frac{R}{Ng}, \quad r_F = \frac{R}{F}, \quad r_{Fg} = \frac{R}{Fg} \tag{4.1}$$

$$N = DCR \cdot T_p, \quad N_g = DCR \cdot T_g \tag{4.2}$$

$$F = \frac{I_0}{\tau}(1 - e^{\frac{-T_p}{\tau}}) \approx \frac{I_0}{\tau}, \quad F_g = \frac{I_0}{\tau}(1 - e^{\frac{-T_g}{\tau}}) \tag{4.3}$$

$$\frac{r_{Ng}}{r_N} = \frac{T_p}{T_g} \tag{4.4}$$

$$\frac{r_{Fg}}{r_F} \approx (1 - e^{\frac{-T_g}{\tau}})^{-1} \tag{4.5}$$

The application of Raman spectroscopy with Ra-I was not studied in detail, but a proof of concept experiment was performed to demonstrate the effect of time-gating on the Raman to noise ratio, indicating the potential of time-resolved Raman using SPAD line sensors with high time-gate resolution. The sample used in this experiment was toluene, providing a strong Raman signal.

4.2 Methods

4.2.1 The optical system

The optical system consists of two main blocks, an epi-fluorescence arrangement and a custom spectrograph (Krstajić *et al.*, 2012; Kufcsák *et al.*, 2017) using a volume phase holographic (VPH) grating and the Ra-I sensor (Fig. 4.4).

In the epi-fluorescence setup, the sample is excited along the same optical path, through the objective lens, as the one used for collection of the emitted light from the sample. This needs specific filters and a dichroic beamsplitter to be used, depending on the wavelengths of the excitation and the expected fluorescence (or Raman) signal. The dichroic beamsplitter allows to direct the excitation light to the sample, through the objective. When light from the sample arrives at the dichroic, it filters it by directing back-reflected excitation light towards the source, and letting light of longer wavelengths (e.g. fluorescent and Raman-scattered light) towards the sensor. The lens was a x10 microscope objective (RMS10X, Thorlabs Ltd, UK), while different dichroic mirrors and filters were used in different experiments. A summary of the configuration for each experiment is given in Table 4.1.

The epi-fluorescence setup is connected to the spectrograph through a fibre patch. Fibre patches of 0.22 numerical aperture (NA) and different core diameters were used in different experiments, depending on the required spectral resolution and optical signal strength. The spectrograph consists of a collimating lens (e.g. of 50mm focal length (AC254-050-A-ML, Thorlabs Ltd, UK), a VPH grating (600 $\frac{\text{lines}}{\text{mm}}$ at 600 nm, Wasatch Photonics, USA) and a focusing lens (e.g. of 75 mm focal length (AC508-075-A-ML, Thorlabs Ltd, UK)). The above configuration gives a spectral range of 200 nm on the 256 pixels of the Ra-I sensor, however, the collimating and focusing lenses can be changed freely to adapt to the required spectral range and resolution. The required spectral range of each sample was determined by measuring its emitted spectrum with a commercial spectrometer (USB2000, Ocean Optics), connected to the output of the epi-fluorescence setup. The lenses of the spectrometer setup were chosen accordingly, to provide a similar spectral range falling onto the line sensor. Wavelength calibration of the spectrometer setup, i.e. mapping the pixels to their detected wavelengths, was carried out by pointing the fibre patch towards a fluorescent lamp, and detect its spectrum on Ra-I. Known characteristic peaks of the lamp spectrum were used to estimate the detected wavelength at each pixel, assuming an even sampling of wavelengths over the pixels. Closely spaced peaks of the spectrum (which could still be resolved) gave indication of the spectral resolution. Resolution measured with this technique was at least 2.5 nm. Wavelength mapping could also be performed with laser light of different wavelengths (e.g. using a supercontinuum

laser with a tunable filter, such as the one described below). The efficiency of the spectrograph was measured to be 68 % at 520 nm. For this, a supercontinuum laser (WhiteLaseMicro, NKT Photonics-Fianium, UK) with a tunable filter (LLTF Contrast, NKT Photonics, UK) was used as a source. The optical power was measured at the input of the spectrograph and after its focusing lens with a power meter (PM100D with an S130C photodiode sensor, Thorlabs Ltd, UK).

	Chlorophyll *a* fluorescence	Thrombin probe fluorescence	Lung tissue fluorescence	Toluene Raman
Focusing lens focal length [mm]	75	100	50	50
Collimating lens focal length [mm]	50	50	30	30
Spectral range [nm]	200	130	180	180
Fibre patch core diameter [µm]	50	105	105	50
Laser	Hamamatsu 402 nm	Hamamatsu 483 nm	Fianium SC <700 nm	Hamamatsu 483 nm
Power on sample [µW]	0.4	0.27	1900	0.68
Dichroic mirror	Semrock FF414	Chroma (custom)	Chroma (custom)	Chroma (custom)

Table 4.1: Summary of the optical configurations used in different experiments. SC - supercontinuum

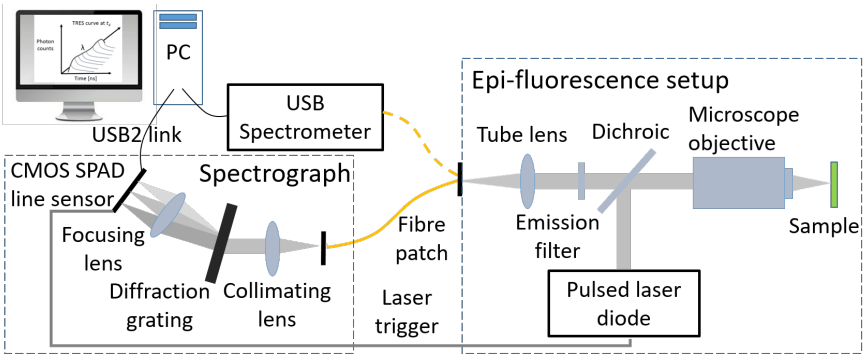

Figure 4.4: The optical system for time-resolved fluorescence spectroscopy consisting of an epi-fluorescence setup and a spectrograph. Revised from Kufcsák *et al.* (2017, Fig. 5).

4.2.2 The Ra-I sensor unit

Printed circuit board

For interfacing with the Ra-I (and Ra-II) sensor, a custom printed circuit board (PCB) was designed by Dr. Ahmet Erdogan, based on a previous PCB design of Dr. David Tyndall (Fig. 4.5). The most important blocks of the PCB are digital-to-analogue converters (DACs), memory modules, a pulse converter circuit, an extension board socket, and several test pins for evaluation of the sensor and PCB. The PCB further allows a high degree of reconfigurability through different jumper configurations.

The extension board socket is used to mount a field-programmable gate array (FPGA) integration module to the PCB. These modules comprise an FPGA for fully customisable hardware designs, to communicate with the sensor and perform data processing (this custom hardware is denoted as firmware throughout the thesis). The extension board used for the experiments reported in this chapter was a XEM3050 integration module (Opal Kelly, USA), encompassing a Spartan-3 FPGA (Xilinx Inc, USA). The extension board provides data connection towards a host computer through a USB link, and related low level software libraries and FPGA modules for fast prototyping. The connection towards the PC in the case of the XEM3050 board is a USB2 link.

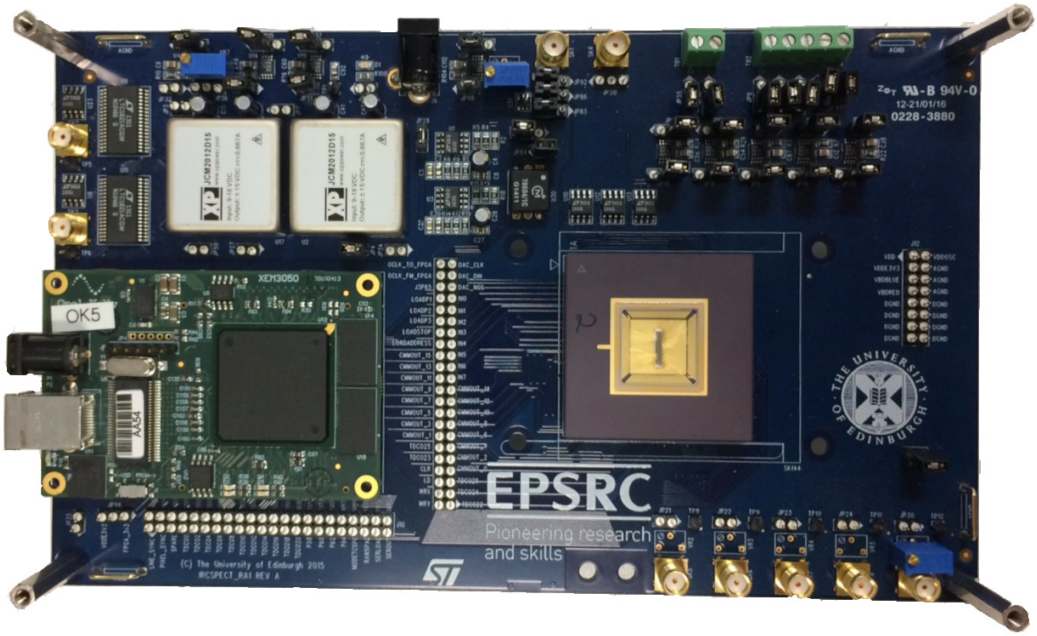

Figure 4.5: The printed circuit board (PCB) interfacing with the Ra-I sensor, designed by Dr. Ahmet Erdogan. The PCB incorporates and extension board with a Spartan-3 field-programmable gate array for fully custom hardware designs controlling the sensor.

The PCB supports synchronized operation between the sensor and the light source.

Depending on the required configuration, the firmware on the FPGA (and the connected sensor) can either act as a slave or a master. This is supported by electric outputs and inputs on the PCB for sync signals. For the latter case, the output sync signal can be generated by the firmware. In a slave mode of operation an external sync signal can be connected to the board. A pulse converter circuit, including a bus translator and an inverter unit, allows to accept different signalling protocols, such as the ones used with nuclear instrument modules (NIMs) or transistor-transistor logic (TTL).

Several different voltages required by the SPAD sensor (for the SPADs and TDC modules) can either be provided from external sources or generated on the PCB, using the DACs. The digital codes for voltage generation are received from the firmware module. Supplementary memory modules allow to generate two independent and custom waveforms, using dedicated DACs on the PCB. These custom signals can be used to drive galvanometer scanners, synchronized to the sensor or the light source. Loading the waveform to the memory modules and addressing them are carried out by the firmware.

Firmware

A custom hardware was realised on the FPGA directly communicating with the sensor, based on previous designs used in the research group, and written using the Verilog language. Specifically, the firmware of a previous line sensor was used as a base, designed by Dr. Nikola Krstajić. This firmware was extended and modified for using with Ra-I, with the data pipeline almost entirely re-written. The primary function of this hardware is to provide seamless control of the sensor for the user, through software calls from the host computer. Hence, this control hardware is referred to as firmware throughout the thesis. The firmware provides data transmission from the sensor to the host computer and performs basic data manipulation tasks. Figure 4.6 depicts the main modules of the firmware architecture. Several modules and preconfigured logic functions (intellectual property (IP) blocks or IP cores) are provided by the software development kit (SDK), supplied by the manufacturer of the extension board (FrontPanel SDK, Opal Kelly, USA), these are shown with blue colour.

Parameters of the firmware are passed from the front end software through the Opal Kelly host interface, and are available in the firmware as wire end points. These include parameters for the sensor operation, DAC codes for the required voltages and parameters determining the dynamics of data transfer. Several trigger signals from the host computer are also used, e.g. for loading firmware parameters into their respective registers and to reset sub-modules. Signals, directly controlling the sensor (such as resetting photon counters, latching the counter values to pixel outputs, addressing pixels and read their data) are handled by the sensor interface module. The same

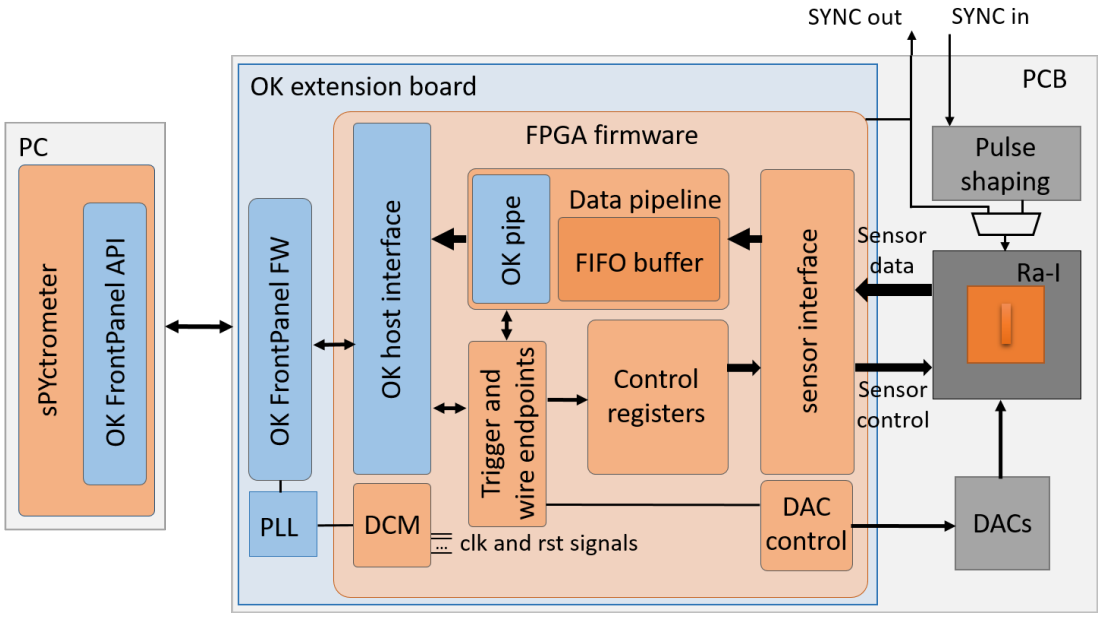

Figure 4.6: Block diagram of the sensor controlling firmware. The firmware forms a layer between the sensor and the front-end software, and takes care of generating the sensor controlling signals and data transmission. DAC - digital-to-analogue converter, PCB - printed circuit board, FPGA - field-programmable gate array, FIFO - first in first out, OK - Opal Kelly, DCM - digital clock manager, PLL - phase-locked loop, FW - firmware, API - application programming interface

module accepts data from the sensor and passes it on to the data pipeline module, synchronously to the sensor control, pixel by pixel. Two finite-state machines (FSMs) in the sensor interface module control the exposure times, and the data readouts from the sensor. As the state-machines can work independently, the user has the option of either performing subsequent or overlapping exposure and read cycles (Fig. 4.7). While the latter case eliminates the overhead of readout between exposure cycles and hence increases the achievable line rate, it demands a minimum exposure time that should not be shorter than the time that reading a line requires.

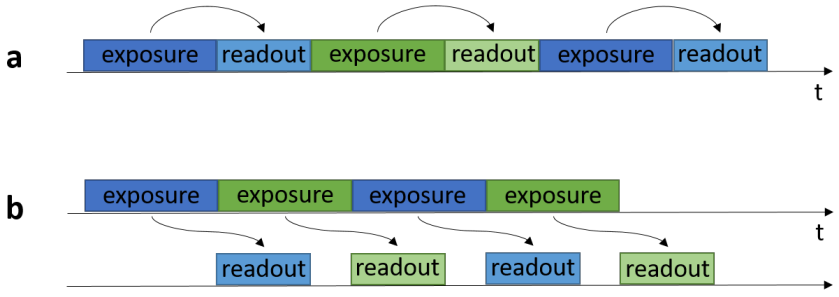

Figure 4.7: Two modes of data readout are offered by the firmware. Subsequent exposures and readouts (a) have no limitations on the exposure times, while overlapping exposures and readouts (b) allow higher line rates.

The data pipeline provides data transmission between the sensor and the host computer. The data pipeline comprises a first in first out (FIFO) buffer to bridge the strict timing of the sensor and the asynchronous transfer of the USB port. The state machines of the sensor interface continuously try to write data to the FIFO buffer while running. On the read side of the buffer, the user can change data transmission dynamics by setting a buffer level at which reading is enabled from FIFO to computer, through a pipe IP block. As a result of the described data transmission method, the sensor data gets lost if reading from the front-end software is not fast enough (i.e. when the FIFO buffer gets full). Since different modules along the data path have to use clock signals of different rates, several clock signals are generated, using a digital clock manager (DCM) IP block. The DCM receives an input clock signal of 32 MHz from a phase-locked loop (PLL) on the Opal Kelly extension board. The clock domains along the data path and the data bus widths between blocks are shown in Fig. 4.8. The highest line-rate between the sensor and the computer was measured to be 19 000 $\frac{\text{lines}}{\text{s}}$, which was limited by the USB2 link (Kufcsák *et al.*, 2017).

Two options support the restoration of lines on the host computer from the stream of raw pixel data. In the first method, the data of the first pixel of each line is marked with a bit pattern. Alternatively, a header can be placed into the data stream just before the data of the first pixel of lines. This header has the same data size as that of a single pixel, and contains three 16 bit coordinates. The coordinates are generated by a separate firmware module, using external electric signals connected to the PCB. These coordinates can be used to assign the recorded lines to three coordinates of a scanning system (e.g. *x*, *y*, *z* spatial coordinates of the field of view, or *pixel*, *line* and *frame* data points of frame-based videos).

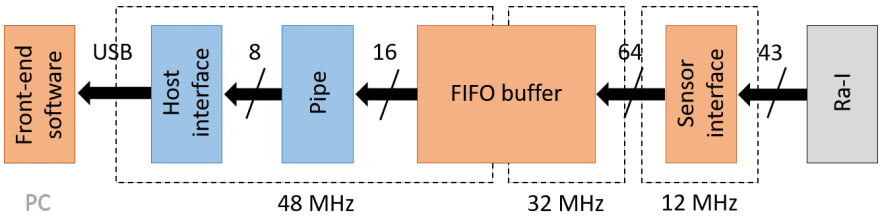

Figure 4.8: Data path from the sensor to the front-end software through the firmware, depicting domains of different clock rates. To bridge the strict timing of the sensor and the asynchronous USB connection a first in first out (FIFO) buffer is used.

Software application

A custom software (sPYctrometer) was written using the Python language (version 2.7, `http://www.python.org`) and the Matplotlib (Hunter, 2007) and NumPy (Oliphant, 2015) packages, providing comprehensive functionality to interact with the Ra-I sensor, the firmware and the PCB. This program is realised in an object-oriented fashion, building on the Opal Kelly FrontPanel library which implements low level routines. The software allows the user to build scripts with efficient functions, performing various tasks with the sensor system, and to design custom software for specific measurement purposes.

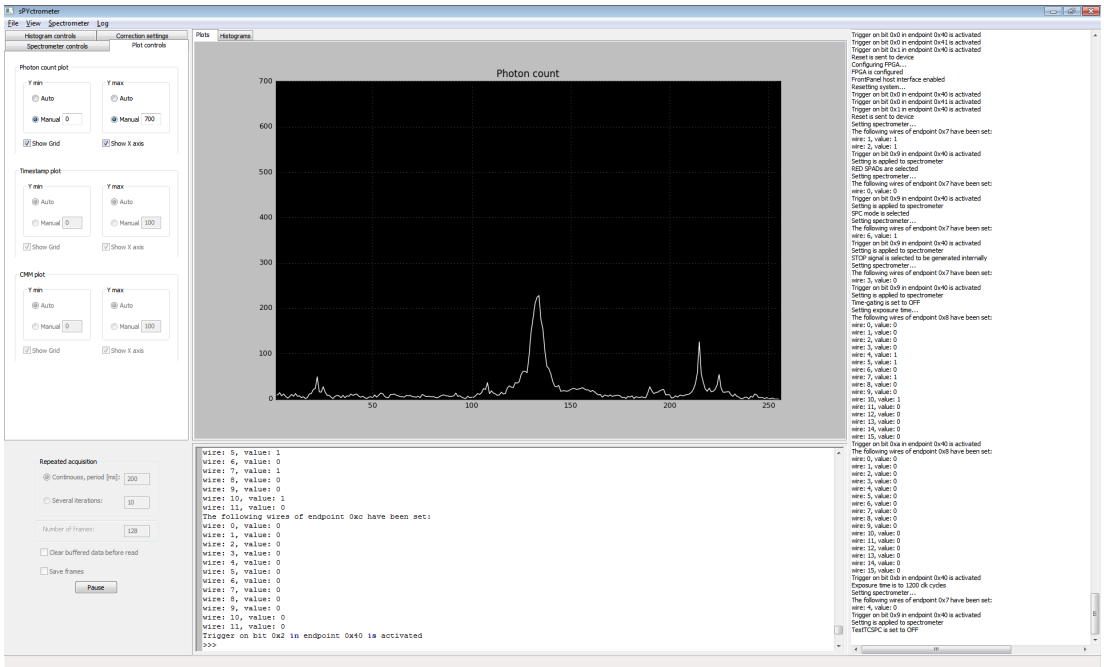

Figure 4.9: Graphical interface for controlling the Ra-I sensor with its firmware and printed circuit board. The software provides access for various parameters of the hardware and firmware system through widgets, and/or user defined scripts through an embedded Python shell. Furthermore, it offers live plotting of the acquired data, and various saving options of the raw and processed data.

A graphical user interface (GUI) is also provided for the end users, utilising the wxPython toolkit (`https://wxpython.org/`), the Python version of the open source wxWidget cross-platform GUI library (`http://www.wxwidgets.org/`). The aim of the GUI is to provide a general interface and easy access to the sensor for researchers at various fields, with potential prototyping and low-volume production of the sensor platform in mind. The GUI allows control of the sensor and its peripheries in a straightforward and uncomplicated manner by hiding fine details, and offers various options for plotting the sensor data. To support experiments, it allows automatic saving of the acquired data and logging all activity performed. The GUI incorporates an interactive Python shell as

well, to offer access to the whole platform and to support specific user requirements. The interface performs data acquisition, computational and graphical tasks (e.g. building histograms and plotting) on a single thread, that is, sequentially. In most cases, when relatively short exposures are used and computationally lightweight tasks are performed, the sequential execution do not limit the user experience. Planned future improvements include multithreaded computation. The GUI and software was used extensively in studies by Ehrlich *et al.* (2017a,b, 2018).

4.2.3 Fluorescence kinetics of chlorophyll *a*

To observe the kinetics of chlorophyll fluorescence as the photosynthesis process is initiated, an intact leaf of *Viburnum rhytidophyllum* was used (picked in December 2016, Edinburgh UK) and placed to the focal point of the objective lens. The leaf was dark adapted by being kept in a closed box at room temperature for at least an hour before starting the experiment. A pulsed laser (PLP10, Hamamatsu, Japan) of 402 nm wavelength and 20 MHz repetition rate provided excitation of fluorescence. The beam was collimated using the lens of the laser head, and directed to the centre of the dichroic mirror. The measured average power falling onto the leaf was 0.4 mW, which relates to 20 pJ pulse energy and 370.4 mW peak power, given the 20 MHz repetition rate, the pulse length of 54 ps, and using the approximation of square shaped pulses. The fluorescence signal and the excitation pulses were separated using a dichroic filter with an edge wavelength of 414 nm (FF414-DI01-25X36, Semrock, USA). The epi-fluorescence setup and the spectrograph were connected with a fibre patch of 50 µm core diameter (M14L01, Thorlabs Ltd, UK). The detected spectral range on the line sensor was 200 nm, provided by the 50 mm collimating lens and 75 mm focusing lens in the spectrometer setup.

Temporal information of fluorescence on each of the pixels of the Ra-I sensor was acquired both in TCSPC and the CMM. These two modes were used in separate experiments, although the sensor system allows changing of the working modes during the same experiment, i.e. one could use the CMM for fast sampling of lifetime changes during the rapid rise of chlorophyll fluorescence and then switch to TCSPC to record more detailed decay curves, but with lower sampling of the reaction kinetics during the slow fall. The timepoint of each spectrum is determined as well, with respect to the start of the experiment.

In the experiment performed in TCSPC mode, an exposure time of 8.3 µs was applied, followed by a sensor readout of 42.9 µs (subsequent exposures and readouts, Fig. 4.7a). The number of detected photons at each pixel was less then 5 % of the number of excitation laser pulses, in order to avoid distortions along the time axis (photon pile-up). Also, the number of detected counts per pixel was less then 50 % of

the number of lines, in order to avoid distortions along the spectral axis (counting loss). 1000 subsequent lines were used to build TRFS cubes, covering an overall exposure time of 8.3 ms over a timespan of about 51.2 ms. The timepoint of the TRFS cubes was defined by the mean timepoint of the lines involved in the cubes. Gaps of few milliseconds between TRFS cubes were introduced by memory management on the host computer. Lifetimes from the TCSPC decay curves were recovered by convolving single exponential decay models with the previously recorded instrument response function (IRF) curves, and fitting to the data using non-linear least square fit routines, based on the DecayFit software (Krstajić *et al.*, 2015; Preus, 2014).

Experiments using the CMM were carried out with an exposure time of 83.3 µs, with overlapping exposures and readouts (Fig. 4.7b). The photon count and integrated timestamp data from the sensor (see section 3.2.1) and the derived centre of mass values were further processed to provide lifetime estimates. Detailed description of the processing steps can be found in Appendix B.

A time-mask of 31.7 ns was used in each of the experiments to increase the SNR of the measured data. This was performed using the counter masking feature of Ra-I (see section 3.2.1) with the edges of the mask signal positioned around the expected photon arrivals (i.e. over the decays). Data acquisition scripts using the sensor controlling functions were customized to suit the dynamics of chlorophyll fluorescence and its fast initial rise and slow fall in intensity. The data acquisition protocol is described in detail in Appendix A.

Intensity information was extracted from TCSPC data by considering only the fact of photon arrivals, without their timing information. For each intensity data point, 1000 consecutive TCSPC lines were binned to reduce statistical uncertainties from low counts, and the data of all sensor pixels were summed. The respective timepoint of each data point was defined as the average timepoint of the 1000 lines used for the data point. Even though this method has significantly lower efficiency compared to single-photon counting (SPC) mode of operation (see section 3.2.1), it allows to collect the intensity profile along with timing information in a single TCSPC experiment. If only the intensity changes are of interest with even higher sampling rates, one would need to use SPC mode with short exposure times.

4.2.4 Fluorescence kinetics of a thrombin FRET probe

In this experiment, fluorescence kinetics of a custom designed FRET biosensor were studied. This biosensor is a probe for detecting thrombin, an enzyme that takes part in blood coagulation. The FRET molecule pairs of the biosensor are 5-carboxyfluorescein (FAM) at one end of the biosensor and methyl red (MR) on the other one, quenching the fluorescence of FAM (Fig. 4.10). Upon induction of thrombin, the substrate of the probe is cleaved and FRET is eliminated, increasing FAM fluorescence. The thrombin FRET biosensor is a patented probe (Bradley *et al.*, 2016) designed by Professor Mark Bradley, Dr. Sunay Vijaykumar Chankeshwara and Dr. Alicia Megia-Fernandez. The solution for the fluorescence kinetics experiment was prepared by Dr. Emma Scholefield.

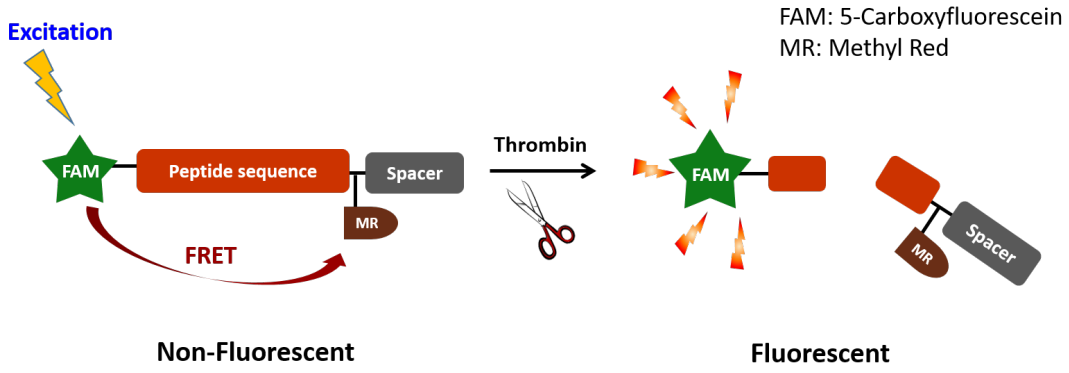

Figure 4.10: Structural changes of the probe molecule causes a change in fluorescence. When the peptide sequence of the biosensor is cleaved by thrombin, 5-carboxyfluorescein (FAM) and methyl red (MR) are separated and FRET is eliminated, leading to an increased fluorescence signal from FAM. Figure from Kufcsák *et al.* (2017).

In the studies with the FRET probe, the same Hamamatsu laser source provided excitation as in the leaf studies, but with a laser diode head emitting at 483 nm. The average power from the laser falling to the sample was 0.27 mW, which equals to 225 mW peak power and 13.5 pJ pulse energy (with a repetition rate of 20 MHz, 60 ps pulse length, approximated with a rectangular pulse shape). The filters in the optical setup (excitation, emission and dichroic filters) were changed to a custom three colour filterset (Chroma, USA) to match the excitation wavelength. The laser light exiting the objective lens was focused into a 5 mm $\times$ 5 mm microcuvette, containing the biosensor in solution.

The biosensor was dissolved in a matrix metalloproteinase (MMP) buffer (10 $\frac{\text{mmol}}{\text{dm}^3}$ CaCl2, 6.1 g Tris-HCl, 8.6 g NaCl per litre, pH 7.5) to a concentration of 5 $\frac{\mu\text{mol}}{\text{dm}^3}$. Approximately 8 s after the start of data acquisition, thrombin (T9326-150UN, Sigma Aldrich,

UK) was added to the solution using a pipette, at a concentration of $5\,\frac{\text{unit}}{\text{ml}}$. The final volume of the solution with thrombin in the cuvette was around 300 µl. For control experiments, a thrombin inhibitor, Anti-thrombin III (AT3, Sigma Aldrich, UK) was added to the solution at a final concentration of $0.4\,\frac{\mu\text{mol}}{\text{dm}^3}$, prior to enzyme activation.

A time-mask of 31.7 ns was used for increased SNR. Data was acquired in a similar fashion as in the leaf studies. 1000 TCSPC lines with 8.3 µs exposure time were used to build TRFS cubes, covering 8.3 ms exposure time over 51.25 ms due to readout cycles between exposures, and a few milliseconds between TRFS cubes introduced by the host computer. Care was taken with respect to the detected light intensity to avoid early photon pile-up and counting loss. More details of data acquisition is provided in Appendix A.

The covered spectral range on the line sensor was 130 nm, given by a collimating lens of 50 mm focal length and a focusing lens of 100 mm focal length. Etaloning fringes (see section 3.4.4) were observed on the acquired TRFS cubes, which were removed by post-processing the data. For this, TRFS cubes were summed along the time axis to retrieve the spectral fluorescence intensity. The etaloning fringe frequency was identified from the power spectrum of the spectral fluorescence intensity. This frequency was used as the bandstop frequency of an infinite impulse response (IIR) filter used to filter the data. Filtering the data corrects for changes in the intensity caused by etaloning, however, timing can be affected by etaloning as well. The structures within the sensor, which cause reflections and eventually etaloning fringes, increase the path of the photons before detection. Since the distance between these structures is in the micrometer range, the extra delay added to the photon arrival times is considered to be negligible, considering the TDC resolution of Ra-I. The added delay due to etaloning, and even more importantly the variation of the added delay might be important with sensors having a higher time resolution. Since the effect is considered to be minimal with Ra-I, single exponential fits were calculated on the original, non-filtered dataset in a similar fashion as in the chlorophyll fluorescence studies.

4.2.5 Auto-fluorescence of excised murine lung tissue

The excitation light for the lung slides was provided by the supercontinuum laser. Components above 700 nm of the laser output were filtered out, resulting in an average power of 1.9 mW falling onto the specimen. In order to receive sufficient amount of fluorescence light on the detector from the sample, the epi-fluorescence and spectrometer setups were connected using a multimode fibre patch of 105 µm (M15L01, Thorlabs Ltd, UK) at the expense of optical spectral resolution. Regarding temporal resolution, the jitter contribution of the fibre is expected to be insignificant compared to that of the detector, and was not taken into account. With sensors of higher time resol-

ution, the timing variation caused by optical components would need to be addressed. The collimating and focusing lenses in the spectrometer were of 30 mm and 50 mm respectively, producing a spectral range of 180 nm on the detector. The filterset used to separate excitation and emission light was the same custom three colour filterset that was used in the experiments with the FRET probe.

Lung tissue of euthanized mice was embedded in paraffin wax using a standard protocol and sectioned on a microtome at 5 µm thickness. After this, the slides were immersed in xylene, 100 %, 95 %, 80 %, 70 % EtOH and dH2O in that order. In order to mimic fluorescent probes emitting in red that are to be detected in the spectral FLIM images, 10 µL of red Inspeck beads 1 % (I-7224, Thermofisher Scientific, USA) were inserted into the sections. The slides were then mounted under a coverslip in water soluble mountant (ProLong Gold, ThermoFisher). In this experiment, the primary aim was to demonstrate the performance of the line sensor, with respect to measuring spectral fluorescence lifetimes for potential spectral FLIM application. The actual lifetime values derived from the measured information are less relevant, in this sense. For this reason, the mountant substance and the coverslip were not tested separately to see if they produce any fluorescence signals, affecting the measured lifetimes. Certainly, this would need to be involved in the analysis pipeline of any commercial or scientific application.

Frames were collected in TCSPC and the CMM as well (not shown). In TCSPC mode, an exposure time of 8.3 µs was used (followed by readout). The light level was sufficiently low to avoid pile-up and counting loss during TCSPC operation. A time-mask of 14.4 ns was used to increase SNR of the measurements. Etaloning was observed on the spectra as with the FRET probe, which was removed in a similar fashion as before, using a digital bandstop IIR filter.

4.2.6 Time-resolved Raman spectroscopy with Ra-I

Raman studies typically require narrow bandwidth laser sources so that the weak Raman-scattered light is confined to a narrow spectral region, for sufficient SNR. In our studies, we used the Hamamatsu driver with 483 nm laser diode head. This light source has a linewidth of 7 nm, which imposes limitations in observing typical Raman signals of biological samples. Still, it is sufficient for demonstrating the sensor performance if a strong Raman-scatterer specimen is used. Low-resolution Raman spectroscopy (LRRS) offers simple and cost effective systems when spectral resolution requirements are more relaxed (Clarke *et al.*, 1999). The sample used in the time-gated Raman studies was toluene in a 10 mm cuvette. The Raman shift of interest belonging to the sample is 3056 $\frac{1}{\text{cm}}$, which produces a Raman signal at ~567 nm for the given excitation wavelength. The repetition rate of the laser was 50 MHz, providing

an illumination of 0.68 mW on the sample. A fibre patch of 50 µm core diameter was used to connect the epi-fluorescence and the spectrometer setups. The focal length of the collimating and focusing lenses of the spectrometer was 30 mm and 50 mm, respectively. The spectral range detected by a single pixel was 0.7 nm. An exposure time of 5 s was used in SPC mode, first without time-masking, then with an 'on' time-mask of 5.6 ns positioned around the time of detecting the laser pulse.

4.3 Results

4.3.1 Fluorescence kinetics of chlorophyll *a*

Photon counts extracted from TCSPC measurements are plotted against the measurement time in Fig. 4.11a. The fast intensity rise took less than 200 ms with the intensity fall over a minute.

Time-resolved spectral cubes at 0.03 s (Fig. 4.11b) during the fast rise, and at 2.04 s (Fig. 4.11c) and 120.04 s (Fig. 4.11d) through the slow fall are shown in Fig. 4.11. The two peaks on the spectra roughly coincide with reports on chlorophyll fluorescence (Franck *et al.*, 2005). A minor increase in spectral lifetimes can be observed between the first two of the displayed TRFS cubes, belonging to the fast rise of the Kautsky-effect. A distinct decrease of intensity and lifetime can be seen between the second two, corresponding to the slow fall. The evolution of time-resolved spectra during the fast rise and slow fall is visualised in supplementary videos of the chlorophyll fluorescence study (Kufcsák *et al.*, 2017).

Decays on a single pixel extracted from the TRFS data are depicted in Fig. 4.12, at several timepoints during the fast rise (Fig. 4.12a) and the slow fall (Fig. 4.12b). Single exponential decay fits confirm the slight lifetime increase during the fast rise (Fig. 4.12c). Decay fits at timepoints during the slow fall reveal an obvious lifetime decrease, indicating fluorescence quenching (Fig. 4.12d).

Finer sampling of the reaction kinetics and less noisy decays during the fast rise are achieved by sacrificing spectral resolution and squeezing TCSPC lines along the spectral axis. Figure 4.13a shows decay curves at several timepoints during the fast rise, each constructed from 150 TCSPC lines with pixels 30 to 175 binned. The remaining pixels at the two sides of the spectrum were discarded as not containing any signal.

Even finer sampling of spectral lifetime dynamics is enabled by the CMM, which provides more efficient photon collection. CMM lifetime estimates of two bands around the 680 nm and 740 nm peaks are depicted in Fig. 4.14, for the first 40 ms of the chlorophyll

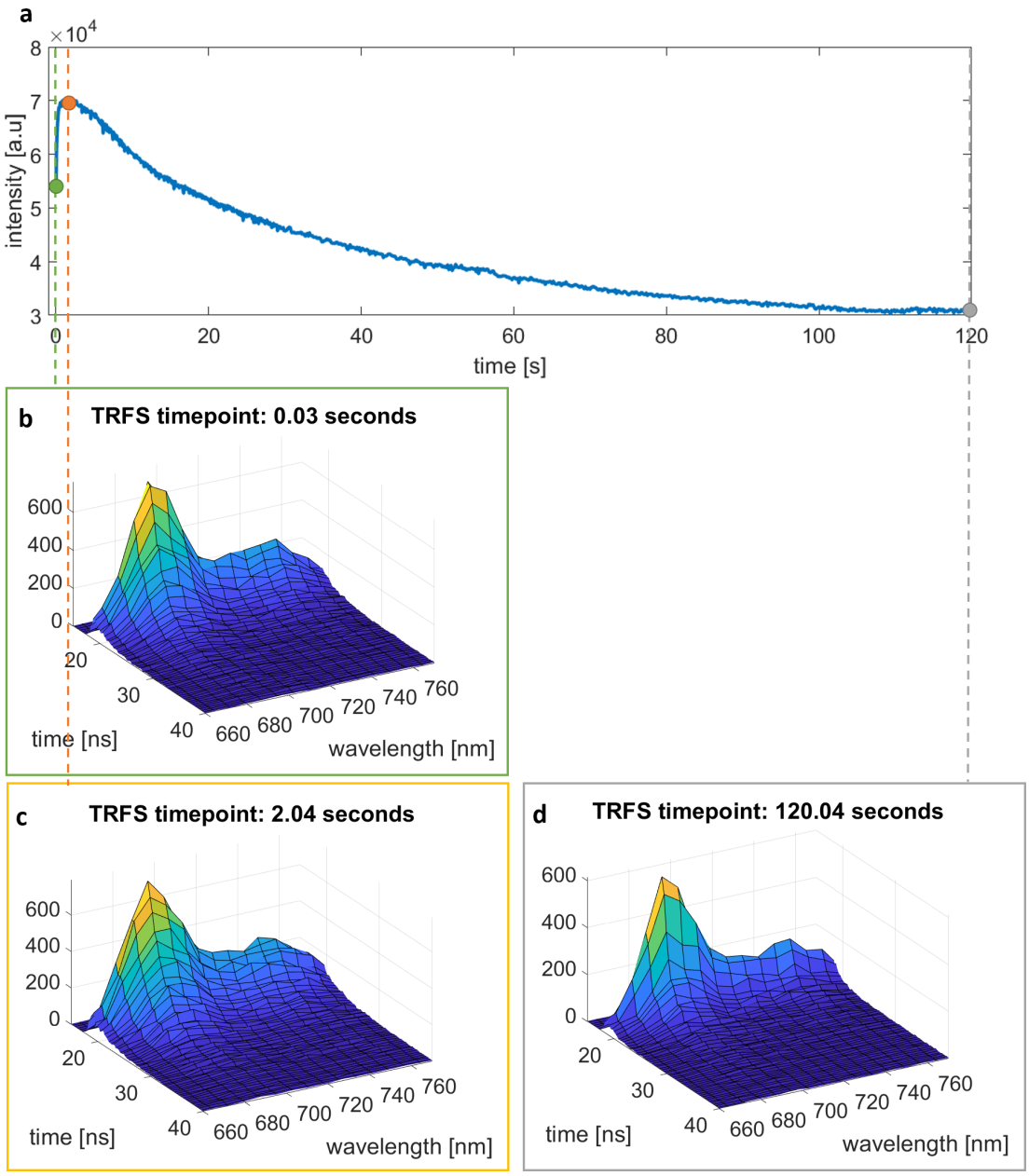

Figure 4.11: Fluorescence intensity transient of chlorophyll *a* over 120 s, extracted from time-correlated single-photon counting measurements (a). Time-resolved spectra at 0.03 s (b), 2.04 s (c) and 120.04 s (d) of the chlorophyll fluorescence transient. The time-resolved plots reveal more information than just the intensity, in this case a change in fluorescence lifetime. The spectral information is also available, showcasing the system's capability of providing a rich dataset for analysing reactions. The time axis was defined using the average TDC resolution of pixels. Spectral bands of 7.7 nm bandwidth were formed by binning every 10 pixel. The spectral content is limited to the band 660 to 770 nm in the plots as no fluorescence was detected outside this band.

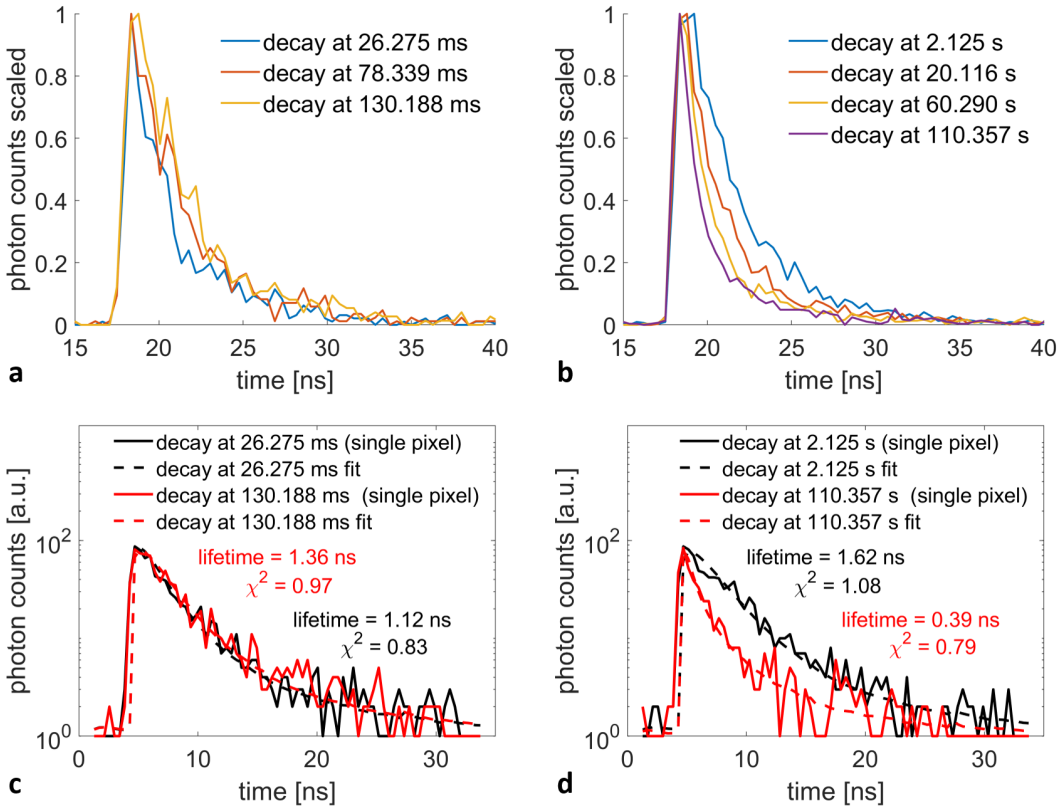

Figure 4.12: Fluorescence decays on a single pixel of 0.77 nm bandwidth during the Kautsky-effect. The decays are extracted from time-resolved spectral histograms of photon arrival times. Decays at 685 nm during the fast rise (a) were taken from a histogram covering an exposure time of 8.3 ms over 51.25 ms. Single exponential fits at 26.275 ms and 130.188 ms timepoints (c) confirm a lifetime increase during the fast rise of chlorophyll fluorescence. The sampling frequency of the reaction during the slow fall is less critical. More accurate decays at 683.5 nm are obtained during the slow fall (b), from arrival times recorded for an overall exposure time of 25 ms over 153.75 ms. The fitted single exponentials at 2.125 s and 110.357 s timepoints show a distinct lifetime reduction during the slow fall. Revised from Kufcsák *et al.* (2017, Fig, 11).

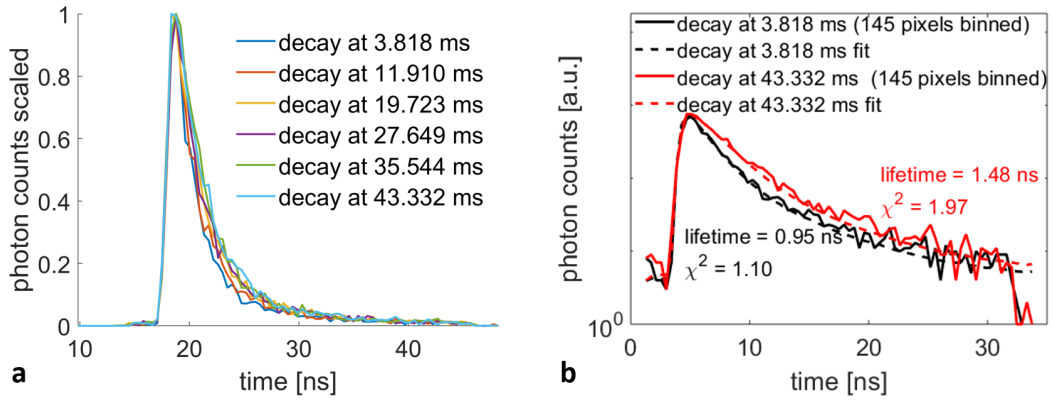

Figure 4.13: Fluorescence decays during the fast rise of chlorophyll fluorescence (a). The decays were built from 150 TCSPC lines of photon arrivals (covering a total of 1.3 ms exposure time over 7.7 ms), binned along the spectrum. Binning all pixels enables finer sampling of the fluorescent transient with less noisy decays, at the expense of spectral information. Single exponential decay fits are calculated for the decays at 3.818 ms and 43.332 ms timepoints (b). Revised from Kufcsák *et al.* (2017, Fig. 12).

fluorescence transient. Merging several CMM lines together provides more accurate lifetime estimates, at the expense of the sampling rate of the transient. Similarly, binning several spectral channels also provides better lifetime estimates, but reduces the spectral resolution.

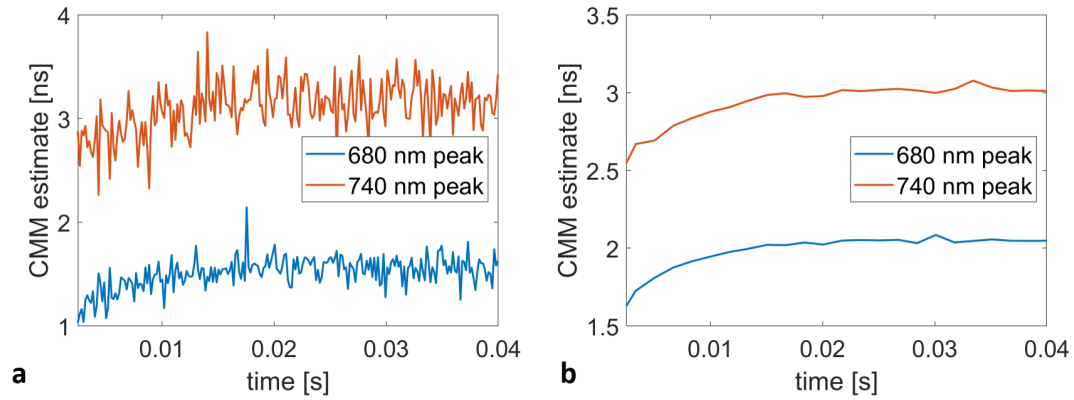

Figure 4.14: Centre of mass of the fluorescence decays over the first 40 ms of the Kautsky-effect, captured using the 'centre of mass' mode of the Ra-I sensor. A high sampling rate of the lifetime dynamics and high spectral resolution is provided when the decay's centre of mass is accumulated over a timespan of 166.6 µs and a spectral range of 0.77 nm (a single pixel) (a). Smoother transients and more accurate estimates of the mean photon arrival is achieved over a timespan of 1.67 ms and a spectral range of 23.1 nm (30 pixels) (b). Figure from Kufcsák *et al.* (2017).

4.3.2 Fluorescence kinetics of a thrombin FRET probe

Fluorescence intensity extracted from the TRFS cubes is plotted against the reaction time in Fig. 4.15. The intensity rise at around 1 s is when the laser was turned on. The dip in intensity around 8 s, when the enzyme was pipetted, is considered to be caused by movement of the fluid in the cuvette.

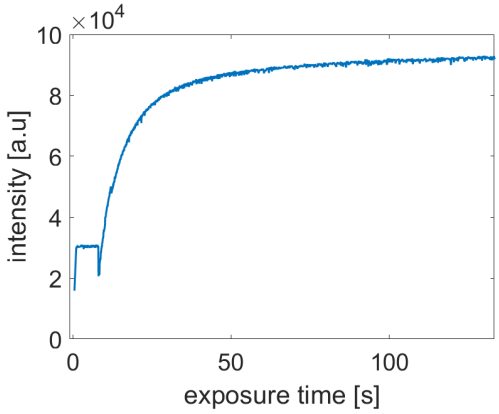

Figure 4.15: Fluorescence kinetics of a Förster resonance energy transfer biosensor for thrombin. Fluorescence intensity starts to rise at 8 s of the experiment when thrombin was added to the solution containing the probe in a cuvette. Intensity values are retrieved from time-correlated single-photon counting measurements. The sampling interval of the reaction was 54 ms in the first 50 s, and 150 ms thereafter. Figure from Kufcsák *et al.* (2017).

A TRFS data cube at 130 s with 1.06 nm resolution is shown in Fig. 4.16. Fluorescence decays, before and after the enzyme activation at 2 s and 130 s, with their respective single exponential fits are depicted in Fig. 4.17. The decays were extracted from single TRFS cubes and 20 pixels binned, resulting in a bandwidth of 10.6 nm. The minute difference in lifetimes before and after the injection of thrombin suggests that unquenched FAM fluorescence dominates the background fluorescence of the solution.

4.3.3 Auto-fluorescence of excised murine lung tissue

A sample TRFS cube of excised lung tissue is depicted in Fig. 4.18. The temporally and spectrally resolved data was generated using 5000 subsequent TCSPC lines, covering a time-span of 257.48 ms, and binning 10 neighbouring pixels (omitting the first 6 pixels of the array), providing a bandwidth of 7.03 nm per spectral channel. Etaloning fringes were removed by digital filtering of the data, for demonstration purposes. A broadband spectrum is visible at 15 ns on the time axis. While this does not affect the measurements and lifetime estimation, it requires further investigation.

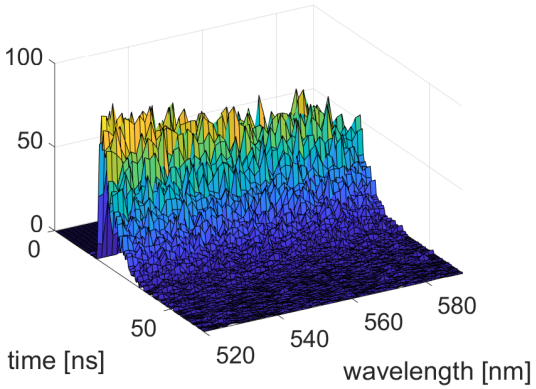

Figure 4.16: Spectral decays at 130 s of the fluorescent transient of a thrombin biosensor. The decays were built from photon arrivals measured over an exposure time of 8.3 ms. The spectral resolution is 1.06 nm (2 pixels). Etaloning fringes were removed using a digital filter. The spectral range is limited to 520 to 590 nm in the plot. Figure from Kufcsák *et al.* (2017).

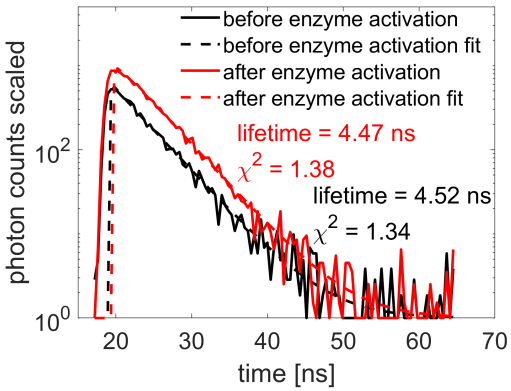

Figure 4.17: Fluorescence decays of a Förster resonance energy transfer based thrombin probe. Decays and related single exponential fits are shown for timepoints before (2 s) and after (130 s) the thrombin enzyme activation. Both decays were built from photon arrivals recorded over 51.25 ms and a spectral bandwidth of 10.6 nm (20 pixels). Revised from Kufcsák *et al.* (2017, Fig. 16).

Figure 4.19a shows decays at 524 nm and 609 nm and their single exponential fits and lifetime estimates, 3.36 ns and 2.12 ns respectively. Shorter lifetimes in the red band (Fig. 4.19b, top curve) are due to the added Inspeck beads, however the high χ^2 values (Fig. 4.19b, bottom curve) suggest an inaccurate model fitted to the data.

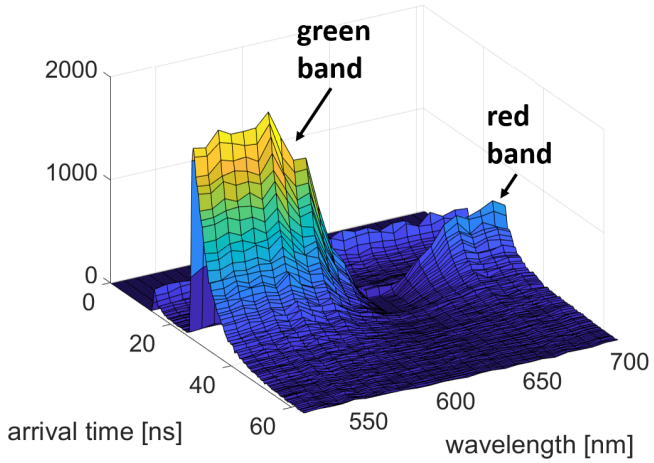

Figure 4.18: Spectral fluorescence decays of excised lung tissue. The decays are built from photon arrivals collected over a spectral band of 7.03 nm (10 pixels) and a timespan of 257.48 ms. No transient of autofluorescence was observed over 50 s of the experiment, as expected. The average time-to-digital converter resolution of 426.8 ps was used to define the time axis. Revised from Kufcsák *et al.* (2017, Fig. 17).

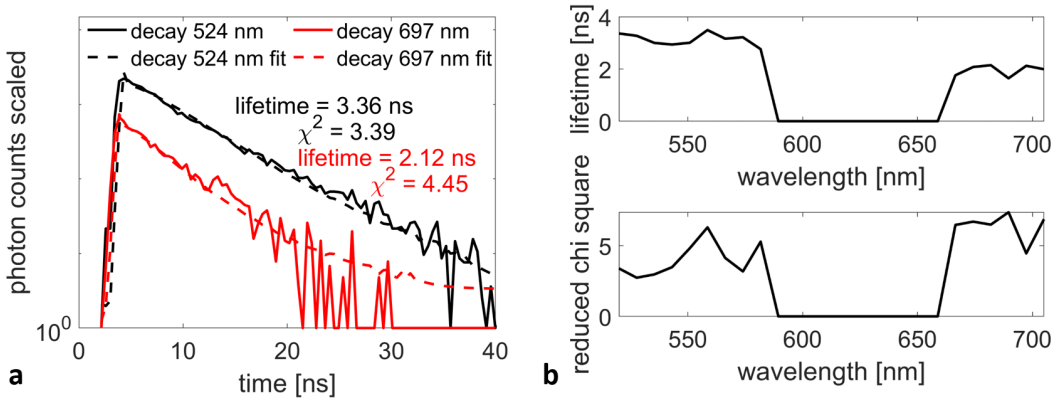

Figure 4.19: Fluorescence decays of excised lung tissue at 524 nm (from green band) and 697 nm (from red band) with 7.03 nm bandwidth (a). Lifetimes in the red band are shorter due to Inspeck beads inserted into the slides with lung tissue (b). Revised from Kufcsák *et al.* (2017, Fig. 18).

4.3.4 Time-resolved Raman spectroscopy with Ra-I

Time-masking performance of Ra-I for time-gated Raman spectroscopy is depicted in Fig. 4.20, showing the Raman spectrum of toluene with and without timegating. As expected, the Raman peak at 3056 $\frac{1}{\text{cm}}$ is quite broad due to the linewidth of the laser (~250 $\frac{1}{\text{cm}}$). As discussed above, the linewidth of the Raman peak does not affect the masking performance. The two curves show a decrease in both the Raman peak and the background noise from DCR. While the latter is an evident consequence of time-masking, the former is an undesirable effect.

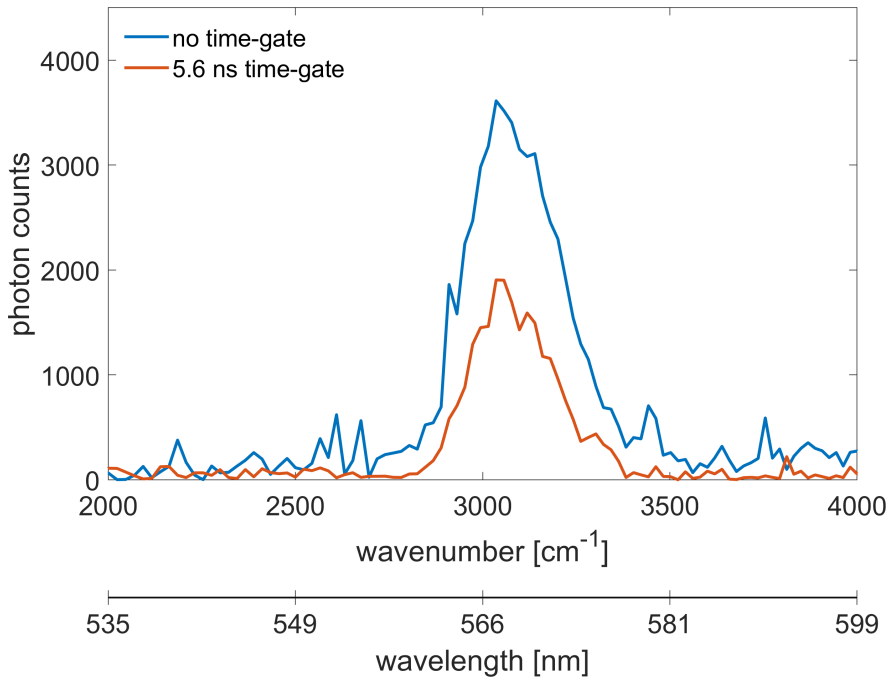

Figure 4.20: Raman spectrum of toluene without time-mask (blue) and with a time-mask of 5.6 ns. The incident optical power on the sample was 0.68 mW and the exposure time was 5 s. Figure from Kufcsák *et al.* (2017).

Quantitative assessment of DCR noise reduction is carried out by calculating the standard deviation of measured intensity values over the range of 2000 to 2500 $\frac{1}{\text{cm}}$, defined as noise. Background suppression is defined as the ratio of the calculated noise values in the non-time-masked and time-masked cases. The estimated background suppression resulted in 3.6, in agreement with the expected value of 3.57, given by the ratio of the laser cycle (20 ns) and the time-gate period (5.6 ns).

4.4 Discussion

The Ra-I sensor and the described optical arrangement, including firmware and software have been demonstrated in several applications as a versatile tool for fluorescence spectroscopy. Our aim was to show that the proposed system is capable of acquiring spectrally and temporally resolved fluorescence data, repeatedly, in short intervals. Two examples where this capability is required are live tracking of spectrally and temporally resolved fluorescence signals from chemical reactions, and spectral FLIM. With respect to the former one, this capability is particularly useful when repeating the reaction under study several times is not feasible, and in point-of-care testing. Video-rate spectral FLIM is advantageous for *in vivo, in situ* medical diagnosis. The two main factors underpinning fast sampling (either temporally or spatially) are the highly efficient processing of photon arrival times, and the high line rate of the sensor.

High photon sorting capability, on one hand, roots from reduced deadtime. Multiple SPADs in a pixel are capable of photon detection even during deadtime of a single SPAD, allowing more events and bursts of photons to be resolved. Time-resolved detection benefits from the multiplexed pixel architecture as well. Provided a considerably shorter TDC conversion time than SPAD deadtime, multiple events can be timed on a pixel during deadtime of a single SPAD. A direct consequence is that the laser cycle can be shorter than the deadtime of a SPAD, allowing higher repetition rates and shorter acquisition times while still maintaining sufficient information for decay fits and lifetime estimation. Most importantly, architectures with several SPADs are capable of detecting count rates well beyond the 1 % pile-up limit, without distortions of the derived lifetimes (Arlt *et al.*, 2013). On the other hand, the multichannel timing architecture further increases the efficiency of photon processing. Assuming that each pixel receives the same amount of light from the sample fluorescence, the maximum number of photons from the sample that can be processed without TCSPC pile-up scales with the number of channels.

The maximum line rate of the described system is 19 000 $\frac{\text{lines}}{\text{s}}$ from sensor to computer, which was the highest reported line rate for a CMOS SPAD line sensor at the time of its publication (Kufcsák *et al.*, 2017). In comparison, this is 27 times higher than the 700 $\frac{\text{lines}}{\text{s}}$ line rate of the predecessor line sensor of Ra-I. This improvement was enabled by a modified PCB (with a wider data bus from sensor to FPGA, designed by Dr. Ahmet Erdogan), and the firmware, described in this chapter. The current line rate is limited by the USB2 connection. In theory, the line rate can be increased with more densely packed information, i.e. concatenating the 47 bit pixel data without padding zeros (see Fig. A.2 in Appendix A). However, this would likely increase the time of processing as it requires intensive bit manipulation of the recorded data. Alternative communication protocols (e.g. gigabit Ethernet) are also to be considered.

4.4.1 Performance in fluorescence spectroscopy

The performance of the optical system with the Ra-I line sensor was demonstrated through three applications in fluorescence spectroscopy. The experiments were chosen to illustrate the versatility of the system, but also to reveal the maximum performance when emphasis is put on either the spectral or the temporal information. By spectral binning, the measurement time window can be shortened for high sampling rate (spatially for FLIM, or temporally for double kinetics). Similarly, one can sacrifice the sampling rate for accurate decays over a wide spectral range. In fact, the system has an inherent flexibility, since the data acquisition is independent of the above choices. Spectral and temporal binning of the measured photon arrival times can be adjusted freely during the analysis of the measured data, with respect to the requirements at different phases of the reaction.

While focusing on speeding up the lifetime measurements either through spectral binning or using less number of photon arrivals, it is essential to consider the minimum amount of information to be collected for accurate lifetime estimates. The requirements for acceptable accuracy (with respect to the variance of the lifetime estimate) can be given as the minimum number of photons (or photon arrival times) to be recorded (Köllner and Wolfrum, 1992). Using the single exponential model (Köllner and Wolfrum, 1992, equation 16) with a lifetime of 3 ns and an observation window of 15.2 ns with 38 time bins (each with 400 ps width), we need to capture at least 120 photons for no more than 10 % error in the estimated lifetime. This calculation is only valid, though, if the applied model is correct (e.g. single exponential distribution with Poisson noise), and the right, unbiased and efficient parameter estimation method is used, such as the weighted least squares or maximum likelihood estimation. A single exponential decay in itself may not be the best model to describe the detected signal, which often has a considerable background and is affected by the sensor jitter. Indeed, for some of the measurements presented in this chapter a more appropriate model would be required. Nonetheless, the above number suggests that the applied acquisition protocol did not limit the attainable accuracy of lifetimes (e.g. the number of photons in the histograms in Fig. 4.13 were above 8000). The accuracy of the CMM can be assessed similarly. For a single exponential decay with no jitter or background, the average arrival time estimates the decay lifetime. Sticking with the single exponential model, the variance of the estimation given by a single photon arrival time can be calculated, again, as given by Köllner and Wolfrum (1992) in equation 16. The variance of the sample mean (provided by the CMM) drops to its fraction by N, with N photons, just as with lifetime estimation through decay fitting. Therefore, using the same number of photons, the CMM provides the same accuracy as lifetime estimation through fitting single exponential models to the TCSPC histograms. However, photons are collected

much faster using the CMM, and hence the same accuracy is reached with shorter acquisition times. Using the CMM becomes challenging with more complex signals. While background and jitter can be corrected for, lifetimes of multiexponential decays cannot be retrieved from the mean photon arrival time. Even in these cases, the CMM value can be used to assess the changes in the temporal profile of optical signals.

As an example of efficient photon collection, decays of chlorophyll fluorescence were observed, during the fast rise of the Kautsky-effect, in a single transient. Previously, this required periodic stimulation (Becker *et al.*, 2014). Commercial instruments provide tens of microseconds timing resolution for measuring the intensity of chlorophyll fluorescence (Stirbet and Govindjee, 2011). In contrast, per pixel histograms were collected in 50 ms with Ra-I (Fig. 4.12a,c). This is much lower sampling resolution, but here we have the lifetime at several wavelengths, not just the intensity. With no spectral information, lifetime values could be measured with roughly 8 ms sampling (Fig. 4.13). As mentioned before, more than 8000 photon arrival times were used in these histograms, suggesting that reasonably accurate decays could still be resolved with a sampling time of a few hundred microseconds (assuming an accurate lifetime estimation with a few hundred photons). At a similar sampling rate (166 µs), the average arrival times (in CMM mode) could be recorded for each of the pixels, preserving spectral information (Fig. 4.14a). At the pixel with the peak of the spectrum (680 nm), the average number of photon arrivals recorded with each CMM line was 800. Assuming that the CMM is accurate using only the third or quarter of this number, we may estimate the fastest data acquisition to be ~50 µs. Sampling the reaction with this time period is reasonably close to what commercial instruments achieve, but in addition, it provides spectral lifetime information, which is not available with those detectors. It should be noted though, that at such short periods the pixels, where the spectral intensity is lower, may not receive enough photons for accurate CMM values. The system with Ra-I is flexible with respect to data processing, the observation window (the time period over which photon arrivals are considered) can be different at each pixel. Accordingly, the sampling rate at these pixels (detecting a lower optical intensity) can be lowered, so that the CMM (or lifetime) estimates are more accurate.

For tracking the thrombin probe activation kinetics, the high sampling rate required by the leaf chlorophyll study is unnecessary. This is especially true, since the spectral content consists of fluorescence of a single fluorophore. Still, resolving complex exponentials over multiple wavelengths with hundreds of milliseconds sampling time period would require high performance of photon processing, for which the proposed sensor system with Ra-I would be highly advantageous.

In the experiments with the lung tissue, decays of spectral bands covering 10 pixels were recorded in 250 ms. If we only use two colours (bands) over the entire detec-

ted spectrum, and assume that in this case the acquisition time drops to 20 ms, a 128×128 image frame would still take 5 minutes to acquire, and the acquisition of an image of 256×256 pixels would take 20 minutes. This is far from state-of-the-art spectral FLIM frame rates (e.g. 8 s per frame of 256×256 pixels, with 64 spectral channels, (Popleteeva *et al.*, 2015)). Experiments in the CMM mode showed that ~1000 counts were processed for a 50 nm bandwidth over 10 µs. Assuming a correct lifetime estimate from 200 detected photons, it is expected that lifetimes with the same spectral bandwidth can be acquired in 2 µs, which is significant in laser scanning confocal systems. The same data rate in a confocal FLIM system equals to about 8 frames per second with frames of 256×256 pixels, and 5 spectral channels. The same frame size with 64 spectral channels would only require 2 to 3 s to acquire. Furthermore, the calculated numbers are based on experiments with 5 µm thick slices of lung tissue, whereas *in vivo* tissue is likely to produce even stronger fluorescence.

4.4.2 Performance in Raman spectroscopy

The Raman experiment showed, that time-masking/time-gating can reduce the noise of the measurement, even though in this experiment the visibility of the Raman peak was not compromised by noise. The amount of noise reduction matches the expectations. However, a drop in the Raman peak intensity was also observed, and therefore the Raman-to-noise improvement is probably less than the noise reduction. Reduction of the Raman peak intensity is most likely to be caused by the long diffusion tail of the red SPADs (see Fig. 3.18) overlapping with the 'off' region of the time-mask. The overlap, furthermore, could be caused by a finite rise time of the gate. To confirm this, time-resolved measurements would need to be performed, where the entire time trace is recorded. This is possible in the TCSPC mode of Ra-I, which, however, has a low photon processing efficiency. It should be noted, that the reason for using time-masked SPC was to maintain efficient photon collection and a reasonably short exposure time. Background suppression of Raman spectra using Ra-I and TCSPC is demonstrated by Ehrlich *et al.* (2017b).

Alternative techniques offer much higher gating accuracy, but have other limitations (Arzhantsev and Maroncelli, 2005). As an example, one could gate the scattered beam through a Kerr gate, and measure the Raman spectrum on a charge-coupled device (CCD) array. While this technique has undoubtedly higher gate resolution and potential for time-gated Raman, direct measurement of fluorescence transient kinetics might be challenging with it, limiting its applicability in multi-modal systems. For spectrally and time resolved fluorescence, the short optical gate (often below a picosecond) can be delayed with various amounts with respect to the excitation laser pulses to sample the fluorescence responses. One of the related challenges is sampling the tail of the decay

(where the fluorescence intensity is low) with high accuracy in a short gate window. If one wants to perform 'one shot' double kinetic measurements, it is also crucial to perform a high number of gated measurements with different delays (for sufficient sampling of the decays), while keeping up with the reaction kinetic. The repetition rate of the laser and the delay unit may constrain this requirement. The laser also needs to provide high energy pulses (for operating the Kerr gate), which raises safety issues in a clinical environment.

4.4.3 Limitations

The flexibility of processing, where the number of counts is increased either through spectral or temporal binning, has limitations. When two fluorophores are observed over the spectrum with different lifetimes, binning the respective photon arrivals would provide and 'average' lifetime. In these cases, fitting a single exponential model to the binned decay may result in a high χ^2 value, hence the decay fitting routine needs to be equipped with the right model. Several other properties of the sensor system infer challenges and limitations, namely time resolution of the TDCs (0.43 ns) and the time-gate/time-mask (1.44 ns), jitter (often > 1 ns) and etaloning.

TDCs of the Ra-I sensor enable the measurement of both fluorescence and the substantially longer phosphorescence decays, given the 27 bits of TDC code with 426 ps resolution (resulting in a total timespan of 57.2 ms). This resolution, however, limits the obtainable lifetimes which are below a ns for biomolecules. High jitter values are also often problematic, especially with time-resolved Raman spectroscopy. The aim here is to separate the fluorescence background from the Raman spectra. Since the spectrum of these two overlap, optical filtering methods cannot differentiate between them. The different temporal response of fluorescence and Raman signals allows their separation in time. However, if the timing accuracy is low, the detector won't be able to tell them apart. Ra-II, a more recent, improved version of the Ra-I sensor yields a TDC bin resolution of ~50 ps and jitter below 200 ps (Erdogan *et al.*, 2019).

The time-gate/time-mask resolution of 1.44 ns is not a major limitation with respect to the position of the gate edge. If one wants to use a time-gate/time-mask closer than 1.44 ns to the signal's time region of interest, an external delay (with higher resolution) is required to be applied to the laser sync signal. However, the gate/mask width cannot be altered to have finer steps and this might limit applications where only a narrow time region of the optical signal is required.

Etaloning fringes required post-processing of the data for accurate intensity measurements. As discussed, the additional delay added to the photon arrival times due to etaloning, and variations in timing are not considered to be relevant, given the

time resolution of Ra-I. The effect is probably more prominent with even more accurate time-resolved sensors, in which case it needs to be characterised. Etaloning is frequently observed with frontside illumination (FSI) sensors with many parallel layers of different refractive index. Backside illumination (BSI) sensors are usually less prone to such effect, although reflections between the front and back surfaces could lead to interference fringes, typically in the near infrared (NIR) region (Waeny, 2014). Other techniques to reduce etaloning include calibration of the detector response to a known excitation, using anti-reflection coating, and special manufacturing processes to reduce the wiring stack above the sensitive area of the sensor or to reduce the amount of parallel surfaces of the sensor structure (IS-Instruments, 2017; Andor). Since no etaloning was observed in the chlorophyll kinetics experiments, it is also possible that the observed fringes were induced by the cover slide and the cuvette, this needs, however, further investigation.

Performance of the entire platform can be increased by improving the firmware and software as well. In particular, more efficient photon sorting would be possible with a higher duty cycle of exposure during data acquisition (that is, faster readout). The data rate from sensor to FPGA is affected by the clock rate of the sensor interface module of the firmware. Experiments with clock rates higher than 12 MHz introduced corrupted data from pixels at one side of the sensor. This is likely due to the propagation time of certain control signals across the sensor chip, which sets the upper limit of the clock frequency of the module. Further investigations would be needed to identify the ideal pulse widths and timings between the limiting control signals in order to further increase the data rate from sensor to FPGA. The occurrence of gaps in the data acquisition (see Appendix A) depend on the host computer's performance. A bigger FIFO buffer is likely to prevent these gaps when the computer cannot keep up with the data transfer. While the current firmware architecture has no major performance limitations, a standard bus architecture is to be considered in the future for easier development.

On the software side, no limitation on data transfer was observed caused by software when simply acquiring and plotting data (including building of TCSPC histograms). Plans for future improvement involve to investigate options for live processing of the data, e.g. on-the-fly lifetime calculation and plotting. A potential software solution for this in a python environment is the cython programming language and compiler (Behnel *et al.*, 2011) which supports the use of fast and efficient C types and functions compiled in an interpreted Python environment.

4.4.4 Conclusion

In conclusion, the described system applying a CMOS SPAD line sensor offers unique capabilities for spectrally and temporally resolved fluorescence measurements with sub-millisecond sampling rate. Efficient photon processing enabled by massively parallelised TDCs offers fast sampling of luminescence, which can be applied either for tracking changes of spectral fluorescence lifetimes in rapid transients or confocal spectral FLIM. Higher gating performance would allow efficient background rejection in Raman measurements, and eventually multi-model diagnostic tools. The sensor and its surrounding arrangement including firmware and software have paved the path for many studies already (Ehrlich *et al.*, 2017a,b, 2018). The platform is intended to be deployed in further studies of quenching and photobleaching of tissue autofluorescence and bacterial probes, and FLIM for bioimaging.

4.5 Further research using the Ra-I sensor unit

The Ra-I sensor platform has been deployed in other studies in time-resolved fluorescence and Raman spectroscopy.

The Ra-I sensor and software with GUI was used with an early version of the PCB and firmware in a fibre optic based spectroscopic instrument (Ehrlich *et al.*, 2017a). This work demonstrates the benefits of time-resolved fluorescence detection over steady-state measurements, and pH sensing capabilities through fluorescence lifetime measurements. The main limitations of the results presented in this study are the low line rate (700 $\frac{\text{lines}}{\text{s}}$) of sensor and long acquisition times of the measurement data (10 min for 1 s measurement). Current performance of the sensor system exhibits a 27-fold improvement in line rate and live plotting of the acquired data showcasing the advanced firmware and software and upgraded PCB.

The sensor with improved firmware and software was utilised in an optical fibre based sensing instrument for endoscopy (Ehrlich *et al.*, 2017b, 2018) using surface-enhanced Raman spectroscopy (SERS). Functionalised gold nanoshells coated with a molecular probe were deposited on the distal end of the optical fibre. SERS signal from the probe was acquired with Ra-I to measure pH in the local environment of the probe and the fibre tip. Time-resolved detection was used to uncover the SERS signal from the fluorescence background of the probe environment (e.g. tissue autofluorescence) and the intrinsic Raman signal and fluorescence background of the fibre. Spectral TCSPC histograms were post-processed by selecting the time interval of SERS detection and subtracting the fibre background and DCR of the sensor. The proposed method was shown to provide a 19-fold improvement in SNR of the SERS signal with a total

exposure time of 10 s and an average excitation power of 0.8 mW, resulting in a precision of ±0.07 pH unit in pH measurements. The same SNR improvement was achieved for measuring the intrinsic Raman signals of cyclohexane and toluene with an overall exposure time of 300 s.

Chapter 5

Spectral-Domain Optical Coherence Tomography with Ra-I and Ra-II

5.1 Introduction

Spectral-domain (SD) optical coherence tomography (OCT) is firmly embedded industrially and clinically, especially in eye care (Swanson and Fujimoto, 2017). Since the inception of spectral-domain optical coherence tomography (SD OCT) in 2002 by Wojtkowski *et al.* (2002), complementary metal-oxide-semiconductor (CMOS) and charge-coupled device (CCD) based sensors have been used to capture the spectral low-coherence interferometry signatures. The present work investigates the application of photon counting sensor arrays in low-coherence and spectral interferometry. This chapter studies the influence of single-photon avalanche diode (SPAD) line sensors and pulsed illumination on OCT, and presents the imaging performance based on core characterisation. First, the following measures are described to evaluate the performance of OCT systems: sensitivity, dynamic range and saturation, imaging depth, and linearity. After this, important SPAD sensor parameters affecting these measures are analysed. The analysis involves detailed simulations and measurements with the Ra-II sensor. Sections below and the following chapter aim to build a framework for improving SPAD sensors for interferometry and OCT in particular.

5.2 Optical coherence tomography performance measures

The aim is to evaluate SPAD technology in OCT and compare it with other sensor technologies by analysing how changing the sensor parameters affects OCT performance. Performance can be measured by qualitative and quantitative assessment of images acquired which, in turn, can be linked to clinical requirements. In OCT, several measures are used to characterize the system performance, from which the following ones are considered.

5.2.1 Sensitivity

Sensitivity refers to the smallest signal detectable by the OCT system, which relates to the faintest structure along the depth that can still be resolved in the images. This smallest detectable signal can be quantitatively described by its power, which is equal to the process noise variance, i.e. where the signal-to-noise ratio (SNR) is equal to 1. In OCT, sensitivity and SNR are often used interchangeably and usually expressed as the power of a perfect reflection ($P_{perfect\ reflector}$) compared to a minimally detected signal power ($P_{faintest\ signal}$), expressed on a logarithmic scale in decibel units (see equation 5.1). In this chapter, the emphasis is placed on how the SPAD line sensor and its various configurations affect sensitivity of the OCT setup. However, it should be pointed out that the measured sensitivity is also affected by other parameters and components of the setup, as well as the applied technique and tools to measure it (Agrawal *et al.*, 2017).

$$sensitivity = \frac{P_{perfect\ reflector}}{P_{faintest\ signal}} = \frac{P_{perfect\ reflector}}{P_{noise}}[dB] \tag{5.1}$$

5.2.2 Dynamic range and saturation

Dynamic range (DR), while seemingly similar to sensitivity, is distinct in OCT context. It is used to describe the ratio of the highest and lowest detectable signal powers within a single acquisition ($P_{saturation}$ and $P_{faintest\ signal}$ respectively in equation 5.2). Dynamic range is usually lower than the SNR of the system, mostly because limitations of the sensor with respect to the highest signal it can detect without saturation/overloading. In many cases, the limited dynamic range is not a problem, since reflections from tissue are several orders of magnitude lower than that of a perfect reflection. If, however, the detector saturates due to strong reflections, the originally sinusoidal fringe modulation will be clipped and distorted. This introduces harmonics appearing in the Fourier counterpart, and results in the detected structures repeated in the images (Raele *et al.*, 2014). Due to the finite saturation level, there is a trade-off between dynamic range and OCT sensitivity. High sensitivity (and low noise) of the sensor allows to pick

up weak signals and faint reflections from the sample, leading to high sensitivity of the OCT system. However, large input signals will saturate the sensor, due to the high signal gain, limiting the system's dynamic range.

$$DR = \frac{P_{saturation}}{P_{faintest\ signal}} = \frac{P_{saturation}}{P_{noise}}\ [dB] \tag{5.2}$$

5.2.3 Imaging depth

The depth observable in an OCT image is affected by several parameters. First of all, certain wavelengths of light will scatter more (in tissue) than others. When scattering is significant, the amount of optical power backscattered from a certain depth might be too low for detection. Therefore, the excitation wavelength is of paramount importance (Jacques, 2013). Other limitations are also introduced by the optical properties of the OCT system. The numerical aperture (NA) of the focusing lens in the sample arm defines not just the transverse resolution, but the depth of field as well where the probe beam is focused (depth of focus (DOF), see section 2.3.2). In SD OCT the sensor introduces significant limitations as well regarding the observable depth range.

5.2.4 Linearity

In SD OCT, the signal of interest (i.e. the amount of backreflection and backscattering in function of depth) is not measured on a single detection element (as opposed to fluorescence intensity, as an example). Instead, the signal is carried in the spectrum, which needs several detection elements to be used in the measurement. This requires that the detected counts reflect the different intensities falling to the pixels across the spectrum. Non-linear response to the incident optical power leads to distortions in the detected fringe modulation, affecting the measured depth profile (Palchetti *et al.*, 2002). In extreme cases this leads to the appearance of harmonics in the inverse Fourier transform, resulting in repeated, overlapping features in the B-scans (Raele *et al.*, 2014).

In the following sections SPAD sensor parameters are analysed through simulations and measurements, with respect to their influence on the above OCT performance measures.

5.3 SPAD sensor parameters

5.3.1 Exposure time

The exposure time affects both the detected signal quality and the frame rate. For the former, we consider the exposure time dependence of the SNR. SNR in OCT is usually defined as the average squared signal over the noise variance (Kalkman, 2017; de Boer *et al.*, 2017) (see equation 5.3). The most important contributors to the overall noise are the noise of the source (relative intensity noise), noise of the detector (e.g. thermal noise) and shot noise (Sorin and Baney, 1992; Kuo *et al.*, 2013). Shot noise originates from quantum mechanical fluctuations, and sets a quantum limit on the accuracy when measuring optical powers. Assuming that the power from the reference arm is much larger than that from the sample, an optimal reference arm power can be found, where shot noise dominates the overall noise. Another often applied technique to reduce the intensity noise of the source is balanced detection (Kuo *et al.*, 2013). In this differential measurement technique, the difference of the interference signal and the source signal is recorded. In our system with a SPAD sensor, we assume two factors affecting the noise, shot noise of the detected optical signal and dark count noise of the SPADs. Assuming a detected signal with a mean of N counts, the mean squared signal is N^2. Shot noise is modelled with a Poisson process. The noise variance ($\sigma^2_{shot\,noise}$) from statistical fluctuations of the detected counts, is the same as the expected value of the detected counts (N), given by the Poisson distribution. Similarly, the dark count distribution is also assumed to be Poissonian, and the dark count noise is described with the variance of dark counts (σ^2_{DCR}), which equals the mean of the recorded dark counts (N_{DCR}). Both the mean detected counts and mean dark counts scale linearly with the exposure time. From this, it follows that the achievable SNR is proportional to the exposure time (equation 5.3). Therefore, there is a trade-off between high frame rate and high SNR. Another limitation on the optimal exposure time in SD OCT is related to loss in signal power due to fringe washout. In this effect, the varying path length (due to motion of the sample and vibrations in the system) causes phase drifts of the fringes. Long integration of the drifting fringes causes a decrease in modulation depth, and consequently a decrease in SNR (Yun *et al.*, 2004).

$$SNR = \frac{\langle I^2_{signal} \rangle}{\sigma^2_{noise}} = \frac{\langle I^2_{signal} \rangle}{\sigma^2_{shot\,noise} + \sigma^2_{DCR}} = \frac{N^2}{N + N_{DCR}} \propto T_{exposure} \tag{5.3}$$

5.3.2 Bias voltage

The bias voltage of SPADs has to be greater then their breakdown voltage for Geiger-mode operation (see section 1.2.2). The breakdown voltage depends on the SPAD structure, the CMOS technology used and the doping profile. In general, breakdown voltages for a 130 nm CMOS process are in the range of 10 to 20 V (Bronzi *et al.*, 2016). Deep sub-micron CMOS technologies (such as the one used with the Ra sensors (Pellegrini and Rae, 2017)) offer high gate density, short SPAD deadtime and highly integrated systems on a single chip.

As discussed in chapter 3, increasing the bias voltage increases the sensitivity of SPADs to incoming light, and consequently the count rate. An increased count rate is expected to result in a higher signal to shot noise ratio (just as an increased exposure time), however the increasing dark count rate (DCR) has to be considered as well. If DCR scales more than linearly with the bias voltage, at high bias voltages it might become dominant over shot noise, worsening the achievable SNR. Furthermore, strong optical signals may saturate the detector when a high bias voltage is applied.

5.3.3 Quench voltage

SPADs of the Ra sensors are passively quenched (Cova *et al.*, 1996). Quenching time of the Ra-II blue SPADs can be tuned by setting the on-state drain-to-source resistance of the metal-oxide-semiconductor (MOS) transistor at the anode of the SPADs (M_Q in Fig. 5.1). This is influenced by the gate-source voltage (V_{quench}), which can take values from a range allowed by the technology (0.8 to 3.3 V (Pellegrini and Rae, 2017)). Count rates detected by a passively quenched SPAD can be accurately described with a hybrid model (Lee and Gardner, 2000), for which a paralysable radiation detector model (William R. Leo, 1994, section 5.7)) with a given τ_d deadtime gives a good approximation (Eisele *et al.*, 2011). According to this model, the detected count rate (m) in function of the incident count rate (n) can be expressed as:

$$m(n) = n\, e^{-n\, \tau_d} \tag{5.4}$$

From this, the maximum achievable count rate (peak count rate, m_{peak}) is:

$$m_{peak} = \frac{1}{e\, \tau_d} \tag{5.5}$$

Shorter deadtimes allow more photons to be processed on a single SPAD, affecting the dynamic range of the OCT system. Increased afterpulsing probability, as a result of fast quenching, has to be considered on the other hand.

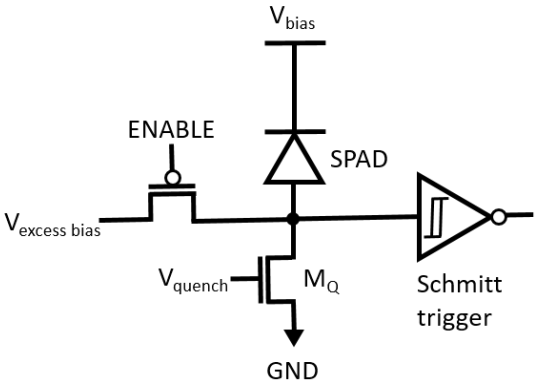

Figure 5.1: Simplified schematic of the blue SPADs of Ra-II. When the SPAD is biased over its breakdown voltage an incident photon triggers an avalanche breakdown. The onset of the avalanche current is sensed on a Schmitt trigger, producing a digital pulse. The quench voltage sets the drain-source resistance of M_Q, affecting the quenching time and effectively the deadtime. When the SPAD is not enabled, the voltage over the SPAD junction is reduced below its breakdown voltage via $V_{\text{excess bias}}$.

5.3.4 Number of pixels

As discussed in section 2.3.4, the spectral range Δk covered by the detector array determines the sampling interval $\delta_s \hat{z}$ of the depth along the sample (see equation 5.6). It was also seen, that the depth of field is limited by the number of pixels (M) covering the Δk spectral range (equation 5.7).

$$\delta_s \hat{z} = \frac{2\pi}{2\Delta k} \tag{5.6}$$

$$\hat{z}_{max} = \frac{M \cdot \delta_s \hat{z}}{2} = \frac{\pi}{2 \cdot \delta_s k} \tag{5.7}$$

Often the number of pixels can be altered, while detecting the same spectral range. This is the case with the Ra-II sensor, where 1024 or 512 pixels can be formed of all SPADs (see section 3.2.2), or when pixel binning is applied on a CCD/CMOS pixel array. Since the detected spectral range (Δk) does not change in these cases, neither does the sampling interval in $\hat{z}$-domain ($\delta_s \hat{z}$). With more pixels, however, the sampling interval of the spectrum ($\delta_s k$) gets shorter, leading to an increase in the total depth in $\hat{z}$-domain ($\hat{z}_{max}$) that is 'seen' by the sensor (recall equation 5.7). The depth increase is obvious when thinking of how denser sampling (i.e. smaller pixels) can sample higher frequency fringe modulations (i.e. higher path differences between the arms). Indeed, the achievable depth range is essentially the folding 'frequency' defined by the closeness of sampling points, which in turn is set by the pixel size.

To recognize this, let us consider a discrete sampling system in general. Half of the sampling rate is called the Nyquist frequency, also known as the folding frequency of the sampling system. Signal amplitudes in the respective Fourier space are symmetrical around the folding frequency. Folding is a consequence of aliasing. Any sinusoidal signal component with a frequency higher then the Nyquist frequency is undersampled. These components have a lower-frequency alias, which produces the same samples in the sampling system. In our sampling system the closeness of sampling points is determined by the pixel size. The density of these sampling points is denoted as spatial frequency, in order not to confuse it with frequencies of the optical signal. The corresponding folding frequency (or Nyquist frequency) is the observable depth range. When an optical signal origins from deeper than the depth range, the corresponding fringes on the spectrum have a modulation frequency that is higher than half of the spatial frequency of sampling. Due to undersampling, aliasing occurs and these fringe modulations are misinterpreted in the depth profile as originating from depths shorter then the depth range.

It follows from the above, that there is a trade-off between the achievable depth range and the axial sampling resolution. The trade-off is set by the number of pixels and the spectrometer optics, determining the dispersion of the optical signal.

The number of pixels affects not just the imaging depth, but the sensitivity and the dynamic range as well. When an optical spectrum with a certain power of illumination is detected on more pixels, each pixel receives a lower optical signal, i.e. fewer counts. Doubling the number of pixels halves the counts per pixel. Halving the count rate per pixel leads to a drop of 3 dB in the signal to noise ratio. DCR of the pixels, on the other hand, is also lower, since a configuration with more pixels is accompanied with fewer number of SPADs per pixel (e.g. switching from 512 to 1024 pixels in Ra-II). Finally, since the saturation level is also dependent on the number of pixels, the dynamic range of the system is affected as well.

5.3.5 Number of SPADs per pixel

The pixels of SPAD based sensors are often formed of several SPADs connected together (Mandai and Charbon, 2012). For example, several SPADs can be connected through an 'OR' gate (Gnecchi *et al.*, 2016), as in the case of the Ra sensors. This technique has multiple benefits. By sharing resources amongst SPADs, such as counters and timing blocks, a significant improve in fill factor (FF) can be achieved. Also, a multiplexed pixel is capable of capturing photons even in the reset phase of one of its SPADs. However, subsequent photon detections at different SPADs of a pixel is only possible, if the pulse width of the SPADs is shorter than their deadtime, so that multiple SPAD pulses do not overlap on the 'OR'-ed node. To that end, the width of

SPAD output pulses is shortened first (Braga *et al.*, 2011) using monostable circuits (Fig. 5.2). The peak count rate (m_{peak}) in this case is determined by the pulse width (T_p) of the monostable circuit (see equation 5.8).

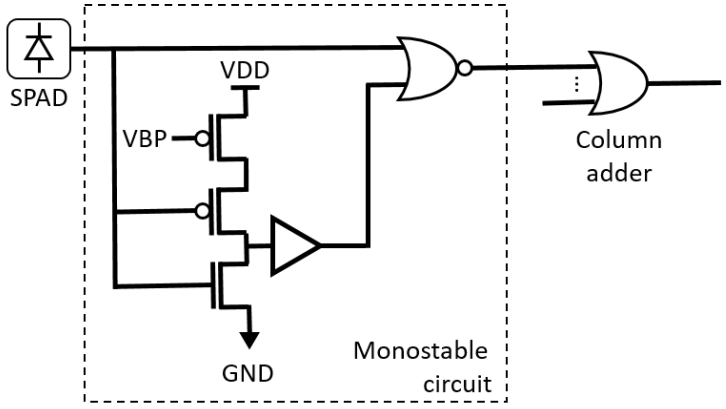

Figure 5.2: The architecture of the monostable circuits of Ra-II for shortening the SPAD pulses. The VBP voltage (0 to 0.7 V) controls the current of the monostable circuit and sets the width of the output pulse (300 ps to 2.6 ns) (Erdogan *et al.*, 2019).

This technique increases the peak count rate and therefore the dynamic range. Due to higher count rates the signal to shot noise ratio is also increased. Unfortunately, the overall SNR is compromised by an increased DCR, as the dark counts of each SPAD are summed as well when 'OR'-ing the SPADs (equation 5.9). Another drawback in time-resolved applications is that the photon timing uncertainty (jitter) increases with the number of SPADs per pixel (Mandai and Charbon, 2012).

$$m_{peak} = \frac{1}{e\, T_p} \tag{5.8}$$

$$DCR = \sum_{i=1}^{M} DCR_i \tag{5.9}$$

The aim is to assess how the SNR in an OCT setup changes with the number of SPADs per pixel. Let us assume a spectrum detected with interferometric fringes on top, caused by a single reflector sample. Let us further assume, that the mean background spectrum and mean DCR are eliminated from the measurement by means of background subtraction. The SNR is calculated as the signal power (signal intensity squared) divided by the noise power (equation 5.10). The signal intensity (I_{signal}) is considered with respect to the amplitude belonging to the reflector on the A-scan. This is directly related to the zero centred interferometric fringes on the background subtracted spectrum. To estimate the expected value of the signal, time averaging is

used (marked by $\langle\rangle$). Let us assume, that the only important noise contribution comes from the statistical fluctuations of the dark counts and the signal counts. The statistical fluctuations are taken into account by the variance of the dark counts (σ^2_{DCR}) and the variance of the signal counts ($\sigma^2_{shot\ noise}$).

$$SNR = \frac{\langle I^2_{signal} \rangle}{\sigma^2_{DCR} + \sigma^2_{shot\ noise}} \tag{5.10}$$

Each of these terms is affected by the number of SPADs per pixel. For example, let us consider the signal in the numerator of the SNR expression. The amplitude on the A-scan belonging to the reflector is proportional to the modulation depth of the respective fringes. The modulation depth increases with the sensitivity of the pixel, which is higher when there are more SPADs in the pixel. Sensitivity in this sense expresses the amount of processed counts in function of the incident counts on the pixel. This is depicted in Fig. 5.3a. Fringes of the incident optical signal (depicted as a sinusoidal curve in function of wavenumbers in the bottom right quadrant) have a higher depth of modulation after detection (red curve in the top left quadrant) when detected with higher sensitivity, i.e. a steeper transfer curve from incident to processed counts (red line in top right quadrant), as opposed to a lower sensitivity case (blue line).

Deriving a closed form expression for the SNR in function of the number of SPADs per pixel is not straightforward. This is especially true when the beam profile has a Gaussian shape along the pixel height (Fig. 5.3b and Fig. 5.8). Instead, a model of uniform illumination is used to assess how the mean signal power and the noise variances change as a function of the number of SPADs per pixel.

Let n denote the incident count rate on a single SPAD with a deadtime of τ_d. The detected count rate of the SPAD in function of the incident count rate is expressed in equation 5.11, using a paralysable detector model. Equation 5.12 expresses the detected count rate of a pixel where M SPADs are active, based on equation 5 of Gnecchi *et al.* (2016). Equation 5.12 can be understood by applying the paralysable detector model to the pixel of multiple SPADs. The input count rate of the multiplexed pixel is the sum of the detected count rates on each SPAD (i.e. M times the detected count rate of a single SPAD). Pulses from the SPADs cannot be differentiated if overlapping in time, therefore the 'deadtime' of the detector model is replaced by the pulse width of SPADs (T_p), which is shorter than the SPAD deadtime, owing to the pulse shortening circuits. We use the expression of the pixel count rate (equation 5.12) to estimate how the signal power changes with the number of SPADs. More specifically, we aim to find the derivative of the pixel count rate with respect to the incident count rate (i.e. the steepness of the pixel transfer curve), as we have discussed that the signal amplitude in the resolved depth profile is proportional to this (recall Fig. 5.3a). Substituting the

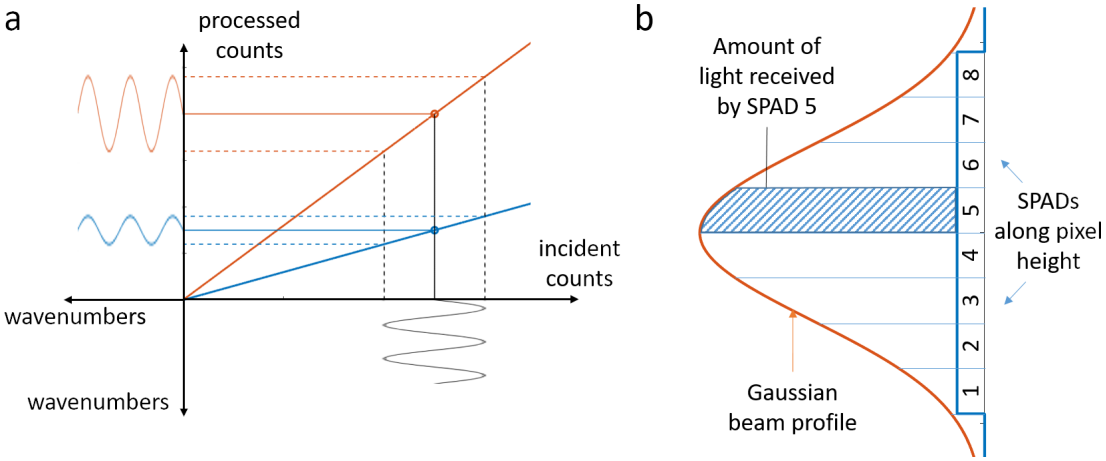

Figure 5.3: The modulation depth of the detected fringes is affected by the slope of the pixel transfer curve (i.e. its sensitivity) which, in turn, depends on the number of SPADs per pixel (a). The dark count rate of a pixel is not related to the modulation depth, but also increases with more SPADs in the pixel. In the case of a beam with Gaussian spatial profile, the amount of signal received is different for each SPAD (b).

SPAD count rate (equation 5.11) in the expression of the pixel count rate (equation 5.12), and calculating the derivative of the pixel count rate ($m_{pixel}(n, M)$) with respect to the incident count rate (n) result in equation 5.13. We see that the expression depends on M, the number of SPADs per pixel.

$$m_{SPAD}(n) = ne^{-n\tau_d} \tag{5.11}$$

$$m_{pixel}(n, M) = M\ m_{SPAD}(n)\ e^{-M\ m_{SPAD}(n)\ T_p} \tag{5.12}$$

$$\frac{d\ m_{pixel}(n, M)}{dn} = Me^{-n\tau_d}e^{-M\ m_{SPAD}(n)\ T_p}(1 - n\tau_d)(1 - M\ m_{SPAD}(n)\ T_p) \tag{5.13}$$

$$DCR_{pixel} = M \cdot DCR_{SPAD} \tag{5.14}$$

It is worth considering how the the pixel count rate and the sensitiveness (steepness of the detected count rate in function of incident count rate, i.e. equation 5.13) change with the number of SPADs, when the incident count rate is low. When the incident count rate is much lower than the reciprocal of the deadtime ($n << \frac{1}{\tau_d}$), the exponential term in equation 5.11 can be ignored. The incident count rate, in this case, is also much lower than the reciprocal of the pulse width ($n << \frac{1}{T_p}$), since the pulse width is shorter than the deadtime. Therefore, the exponential term of equation 5.12 is also ignored, and the pixel count rate is approximated as M times the SPAD count rate. Continuing with the low incident count rate assumption ($n << \frac{1}{\tau_d} < \frac{1}{T_p}$), in the expression of 5.13 both the exponential terms and the last two terms in parenthesis can be approximated as 1. Hence, the derivative of the pixel count rate is approximated to be M (the number

of active SPADs in the pixel).

Lastly, the pixel DCR in function of the number of SPADs per pixel is analysed. The assumption is that each SPAD yields Poisson distributed dark counts around the same median DCR. The pixel DCR is therefore M times the SPAD DCR (equation 5.14).

Summarizing the above results, the signal power is proportional to the steepness of the pixel transfer curve (number of detected counts versus incident counts) squared (equation 5.15). With the approximation of low incident count rates (much lower than the reciprocal of the SPAD deadtime), this is approximated with the second power of the number of active SPADs per pixel (M^2). The amount of shot noise equals the expected number of detected photons, which is given by the pixel count rate multiplied by the exposure time (T_{exp}) (equation 5.16). The pixel count rate is closely proportional to the number of SPADs per pixel (M), given the low incident count rate approximation, as discussed above. Similarly, DCR noise (i.e. the dark count variance) equals the expected number of dark counts, which is calculated as the DCR of the pixel multiplied by the exposure time (equation 5.17). Pixel DCR, and therefore the DCR variance is also approximated to be proportional to the number of active SPADs (M). The final result is that the SNR of the OCT system is proportional to the number active SPADs per pixel (equation 5.18).

$$\langle I^2_{signal} \rangle \propto (\frac{d\, m_{pixel}(n, M)}{dn})^2 \approx M^2 \tag{5.15}$$

$$\sigma^2_{shot\ noise} = m_{pixel}(n, M) \cdot T_{exp} \propto M \tag{5.16}$$

$$\sigma^2_{DCR} = M \cdot DCR_{SPAD} \cdot T_{exp} \propto M \tag{5.17}$$

$$SNR = \frac{\langle I^2_{signal} \rangle}{\sigma^2_{DCR} + \sigma^2_{shot\ noise}} \propto \frac{M^2}{M} = M \tag{5.18}$$

The above approximation is valid up to incident count rates close to the reciprocal of the SPAD deadtime (see also section 5.5.5). Count rates above this should be avoided anyway, due to saturation and linearity issues. These are discussed in a later section. The other extreme, where the approximations start to fail is when the number of SPADs per pixel (M) is approaching the reciprocal of the SPAD count rate times the SPAD pulse width ($\frac{1}{m_{SPAD}\ T_p}$). Practically, this would mean several hundreds of SPADs per pixel, and therefore it is safe to ignore this case as well. The main limitation of the derived result on the SNR is the assumption of a uniform illumination. In a more realistic scenario the incident beam has a Gaussian spatial profile (Fig. 5.3b). In this case, as we activate more and more SPADs of the pixel, the detected count rate increases with the integral of the Gaussian profile (following the sigmoid shape of the Gaussian error function (Andrews, 1997, chapter 3.2)), which after a while, is

lower than a linear increase. In such situation the SNR shall increase initially, when turning on the first few SPADs. However, it is expected that the SNR drops when further increasing the number of active SPADs, as DCR noise increases more heavily, overcoming the diminishing benefits of increased count rate.

SPAD based sensors have been applied in a wide range of application in biophotonics (Bruschini *et al.*, 2019). The benefits of these sensors have not been fully exploited in OCT yet, even though single-photon counting and a SPAD sensor were demonstrated in a time-domain OCT system providing improved sensitivity (Mohan *et al.*, 2008). The aim is to introduce SPAD sensors, for the first time, in spectral-domain OCT. In particular, the timing aspects of SPADs is addressed, and potential applications combining time-correlated single-photon counting (TCSPC) and spectral interferometry. As an example, time-gating was shown to increase the contrast and the depth of field in OCT (Muller and Fraser, 2009). These ideas are introduced in the next chapter. In preparation for this, the primary aim of this chapter is to investigate the feasibility of SPAD line sensor based SD OCT, to reveal the influence of sensor parameters on OCT performance and to identify the related challenges through simulations and measurements.

5.4 Methods

5.4.1 The optical coherence tomography system

The OCT system consists of a broadband light source, a fibre coupler based Michelson-interferometer, and a custom built spectrometer with a sensor connected to a computer for spectral detection of light (Fig. 5.4). In order to evaluate the performance of SPAD sensors for OCT and compare them to the commercially available sensors, the Ra-I and Ra-II SPAD line sensors and a commercial CMOS imager were all tested in the OCT system. Different light sources were also cross compared for OCT, a continuous wave superluminescent diode (SLD) and a pulsed laser source. While the former is suitable for interferometry using photon counting, time-resolved detection, where SPADs are pre-eminent, requires a pulsed source.

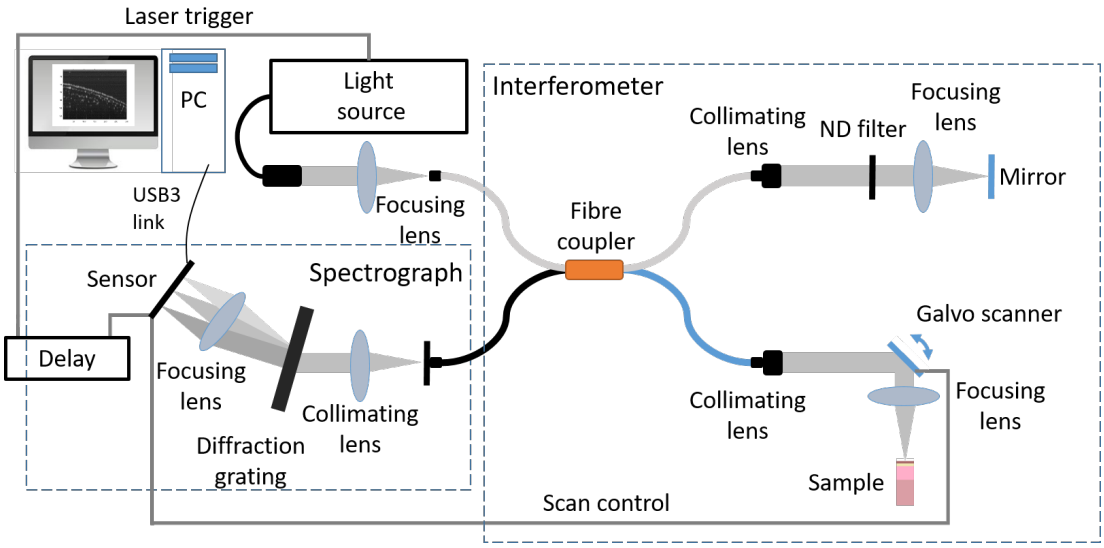

Figure 5.4: Schematic of the system used for the evaluation of CMOS SPAD sensors in OCT applications.

Light sources

The light source in an OCT system has fundamental influence on the imaging performance. Due to high scattering of light in tissue, the illumination has to be in the near infrared (NIR) regime. A large bandwidth is also essential, as in low-coherence interferometry the bandwidth of the source is inversely proportional to the achievable axial (depth) resolution. By definition, the axial resolution δz is the full width at half maximum (FWHM) of the axial point spread function (PSF). This is expressed in equation 5.19, where the source is approximated with a Gaussian shape with λ_0 centre wavelength and $\Delta\lambda$ FWHM bandwidth (Swanson *et al.*, 1992). This equation shows how the axial PSF changes with the bandwidth of the source. When characterising real systems, the assumptions behind equation 5.19 are often too idealistic, resulting in inaccurate resolution estimates. Methods for accurate measurement of the PSF and axial resolution were suggested using resolution test artefacts (Tomlins *et al.*, 2011).

$$\delta z = \frac{2\ln(2)}{\pi}\frac{\lambda_0^2}{\Delta\lambda} \tag{5.19}$$

In the studies for evaluating SPAD performance in OCT, two different light sources were applied and compared. The first one was a source most commonly used for OCT, the SLD with continuous illumination. The second one was a pulsed laser. These light sources provided similar powers and spectral ranges in the experiments.

Superluminescent diode In one case the light source was an SLD (EBS8000 Broadband Light Source, Exalos AG, Switzerland). The spectrum of the source was measured by a commercial spectrometer (Flame, Ocean Optics, USA, Fig. 5.5). Since this spectrometer is only calibrated for the measured wavelengths and not for the spectral light intensities, the measured shape of spectra may differ from their true shape. This makes the estimation of FWHM (the width of the curve at 3 dB points) difficult. As an example, the spectrum depicted in Fig. 5.5 has an FWHM value of 52.76 nm, due to the high peak at around 850 nm. The actual spread of the curve is around 114 nm, and the centre wavelength is at 802 nm. As mentioned above, accurate estimation of the axial resolution would require the use of resolution test targets.

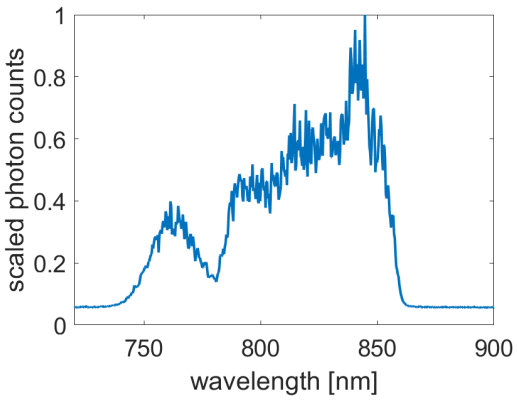

Figure 5.5: Spectrum of the filtered output of the superluminescent diode.

The light was directly coupled into the fibre coupler using the FC/APC output connector of the SLD. The optical power exiting the fibre coupler was measured with a photodiode (S130C with PM100D power meter, Thorlabs Ltd, UK) to be 960 µW at each arm of the interferometer.

Supercontinuum laser One of the main advantages of SPAD based sensors is the capability of time-resolved photon detection with high resolution. To evaluate potential benefits of time-resolved detection in OCT (see chapter 6), a pulsed laser source is required which has a reasonably broad spectral band as well. To that end, a supercontinuum laser was used (WhiteLaseMicro, NKT Photonics-Fianium, UK) with a repetition rate of 20 MHz. The output spectrum of the supercontinuum laser runs from 400 to 2000 nm. This is considerably wider than what is required by the OCT application. Therefore, the spectrum of the laser was restrained by a set of optical filters (FESH0900, FEL0700, FB850-40, Thorlabs Ltd, UK, Fig. 5.6a,c). Even though certain infrared (IR) wavelengths are passed through the filters (Fig. 5.6b), these were confined by the enclosure of the laser and the filter system. No IR light through the fibre coupler was detected either.

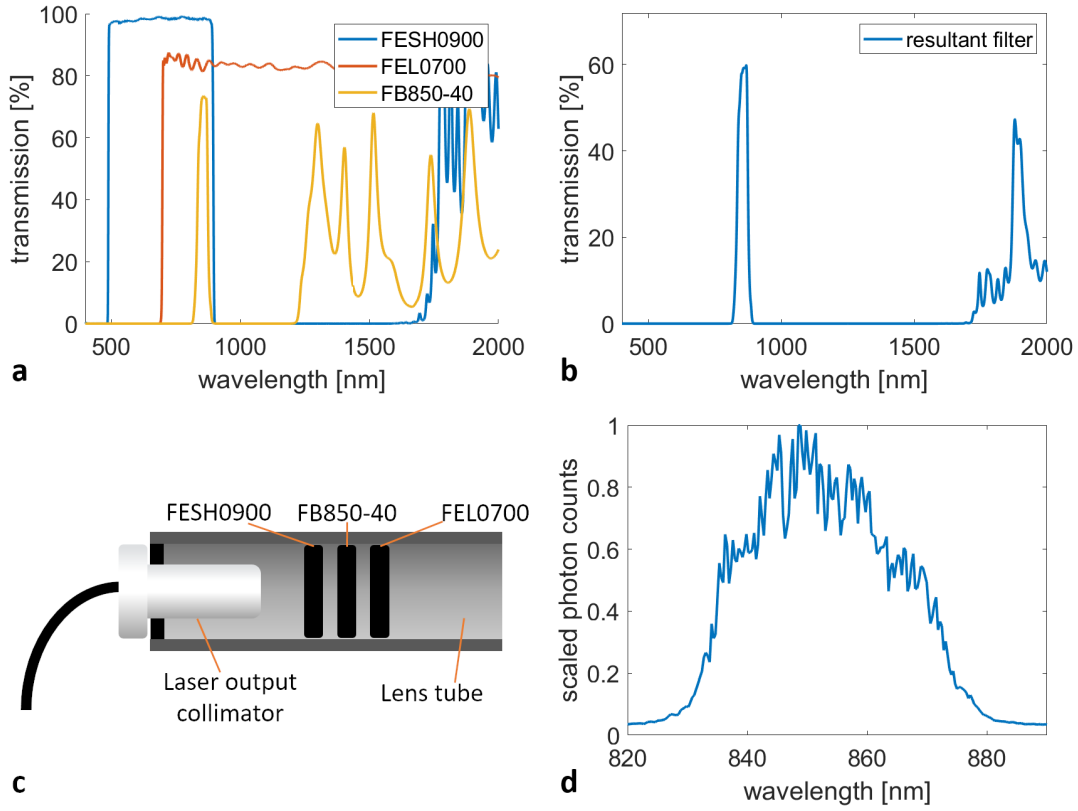

Figure 5.6: Transmission of the optical filters used to reduce the bandwidth of a supercontinuum laser for spectral-domain optical coherence tomography (a). Transmission data is obtained from the manufacturer's website. Transmission of the combined filters (b) has some leakage of infrared (IR) content. This does not introduce any safety hazard, as the laser source and the filters were enclosed and no IR content delivered by the fibre coupler was detected either. For additional safety, IR filtering goggles were used throughout the experiments. The three filters were stacked in a lens tube with the supercontinuum output collimator (c). The output spectrum was confirmed using a commercial spectrometer (d).

The filtered spectrum has a bandwidth of ~38 nm (FWHM) centred at 850 nm (Fig. 5.6d), which was measured using the Ocean Optics spectrometer. A rough estimate of the free-space axial resolution with equation 5.19 is 8.4 µm. The filtered light was coupled into the fibre coupler using an objective lens (10X Nikon Achromatic Finite Conjugate Objective). The optical power exiting the fibre coupler was measured to be 950 µW at each arm of the interferometer.

Interferometer

The light from the source is directed to the sample and reference arms of the interferometer with a ratio of 50:50, using a 5 µm core single mode fibre coupler (FC850-40-50-APC, Thorlabs Ltd, UK). The light exiting the fibre coupler is collimated with fixed focus collimating packages (F280APC-850, Thorlabs Ltd, UK) in both arms of the interferometer.

In the sample arm the collimated beam is diverted with a galvanometric scanner (6210h, Cambridge Technology, USA) in order to perform transversal scanning along the surface of the sample of interest. The beam is focused to a confined spot on the sample by a 30 mm focal length objective (AC254-030-A-ML, Thorlabs Ltd, UK). The focused spot size (δx) is estimated to be 8.1 µm (see equation 5.20, where NA is the numerical aperture of the objective lens and λ_0 is the centre wavelength of the source, (Drexler and Fujimoto, 2015, chapter 2)).The size of this focused spot determines the lateral (transversal) imaging resolution. This theoretical resolution may not be accurate in real applications. The accurate lateral resolution needs to be determined through measurements, just as in the case of the axial resolution. Furthermore, the resolution needs to refer to the capability of resolving two points in the OCT image, close to each other. Equation 5.20 assumes that the intensity values of the two points are linearly scaled, whereas the OCT images are logarithmically scaled. Therefore, resolution is best evaluated when it is based on the OCT images, using test targets (Tomlins *et al.*, 2011).

In most experiments, where different parameters of the sensors were analysed, a mirror was used as a sample. Since no scanning was required in these tests, the scanner was set to a fixed position. In other experiments imaging was required, and the galvanometric scanner was used to move the focused spot along the surface of the sample. The scanner was controlled using electric signals generated by the digital-to-analogue converters (DACs) on the SPAD sensors' printed circuit board (PCB). Interference fringes with high modulation depth requires that the two arms of the interferometer provide similar optical powers. For this reason, neutral density (ND) filters are placed in the reference arm to attenuate the light to a level similar to that of the sample arm. In order to minimise dispersion effects (and the deterioration of the spectral fringes) the reference arm objective matches the sample arm objective.

$$\delta x = 0.37 \cdot \frac{\lambda_0}{NA} \tag{5.20}$$

Spectrometer

For measuring the interference spectrum, a custom spectrometer is built based on the work of Krstajić *et al.* (2012). Backreflected and backscattered light of the two arms is combined in the interferometer. The combined light from the fibre coupler enters the spectrometer, where it is collimated by a lens. The focal length of the collimating lens is determined by the required NA to collect all the light exiting the fibre tip. For this, an achromatic doublet of 50 mm focal length was used (AC254-050-B-ML, Thorlabs Ltd, UK). The collimated light is dispersed by a volume phase holographic (VPH) transmission grating (1200 $\frac{\text{lines}}{\text{mm}}$ at 840 nm, Wasatch Photonics, USA). A second lens is used to focus the collimated and dispersed light onto the detector plain. Longer focal lengths allow more space for the diverging spectral components to spread out. Therefore the spectral range falling on the line sensor decreases. The ratio of the focal lengths of the focusing and collimating lenses, on the other hand, determines the magnification of the two lens system, and hence the spot size of a single wavelength on the detector. With this in mind, the focusing lens was either an achromatic doublet of 75 mm (AC508-075-B-ML, Thorlabs Ltd, UK) or 150 mm (AC254-150-B-ML, Thorlabs Ltd, UK) focal length, depending on which sensor was used. The overall efficiency of the spectrometer setup was measured to be 85 %. This was assessed using a commercial power meter (PM100D with S130C, Thorlabs Ltd, UK). The power of the optical signal was measured at entering the spectrometer, and in focus of its focusing lens.

Sensors

The two SPAD line sensors, Ra-I and Ra-II, were tested in the OCT setup. Several parameters of Ra-II (such as sensitivity, number of SPADs, number of pixels, exposure time and operating voltages) and their effects were evaluated and analysed. In order to cross-compare the efficiency of SPAD sensors with that of traditional sensors used in OCT, a third sensor, which was available in our group, was also tested in the setup. This sensor was a commercial CMOS imager (GS3-U3-41C6NIR-C, FLIR Systems, Inc. - Point Grey Research, Inc., USA) with 2048 $\times$ 2048 pixels of 5.5 µm pitch.

Firmware, software

The firmware described in 4.2.2 was slightly modified for the initial tests with the Ra-I SPAD line sensor. The design was adapted to a Spartan 6 field-programmable gate array (FPGA) (Xilinx Inc, USA) on an XEM6310 FPGA integration module (Opal Kelly, USA), and extended to control the galvanometric scanner. For this, an intellectual property (IP) memory block was synthesized and loaded with a range of increasing digital codes. The digital codes are passed to a DAC module on the PCB, and the

generated analogue voltage is connected to the scanner. The firmware performs a user defined number of exposure cycles, and controls the scanner by addressing subsequent memory entries along with each exposure. By selecting the number of exposures (A-scans) and the step size between memory addresses, the user has full control over the transversal scanning range and resolution. In most experiments 300 scanning steps and a step size of 8 µm were used. In order to minimise the lag between OCT images, the firmware performs bi-directional scanning. The python based software and graphical user interface (GUI) for Ra-I were also extended to process the raw data and plot the OCT images. The achieved frame rate was 2–3 frames per second with 300 scan positions per OCT image frame, which was limited by the exposure time of the A-scans for sufficient SNR.

Firmware of the Ra-II sensor (including the scanning feature) was implemented by Dr. Ahmet Erdogan. Custom matlab scripts were written to read, process and plot the acquired data. The scanner was controlled by the Ra-II firmware and PCB when the CMOS imager was used. In these cases, the Ra-II firmware started the OCT imaging by controlling the scanner. In addition, a synchronous trigger signal was sent to the CMOS imager as well. The trigger signal initiated the acquisition of A-scans on the imager at each scanning position.

5.4.2 Data processing

In order to retrieve the depth profile of the sample of interest, several steps of data manipulation have to be carried out (Fig. 5.7). Each recorded interference spectrum is transformed to a single A-scan, which represents the amount of backscattering and backreflection along the depth at a single transversal scanning point. Several A-scans, one at each scanning point, produce a two dimensional image of the depth structure along the scanning line.

Transformation of spectral lines to A-scans was performed based on the methods described in section 2.3.5. In the first step, a background line was subtracted from the acquired interference spectra. The background subtraction removes the DC component of the fringes and the respective reflection at zero depth on the A-scans. In addition, it serves as a noise reduction step, removing fix pattern noise of the detector. The average background line for background subtraction was calculated from 300 lines, using the reference arm only. The background subtracted data was resampled using cubic spline interpolation. The new sampling points were defined with equidistant spacing between the reciprocal of the wavelengths at the edges of the detected spectrum. After resampling and before the FFT, further filtering was carried out by means of applying a Hann window on the data. This step is performed to prevent spectral leakage, a signal processing artefact. When the measured interfer-

ence spectrum does not diminish to zero at the edges of the detector, the spectral 'window' cuts the edges of the real spectrum. The Fourier transform of the windowed spectrum introduces spurious signal content, and effectively the appearance of false reflections in the depth profile. The aim of a window function is to smooth the edges of the windowed spectrum, and therefore to suppress the spurious reflections. The Hann window was chosen for sufficient stopband rejection, however numerous other window functions exist with their own benefits and drawbacks (Prabhu, 2014, Chapter 5). Fourier transformation was performed by fast Fourier transform (FFT), after which the squared magnitude of the signal was calculated. Finally, the logarithm of the data was taken in order to compress the values covering several orders of magnitude. As an additional step, deconvolution with the optical system's PSF can be carried out to increase the sensitivity of the OCT system (Woolliams *et al.*, 2010). In this chapter, the emphasis was put on revealing the effect of SPAD parameters on OCT performance and finding the best set of parameters. For this reason and due to the complexity of the PSF measurements, no deconvolution was performed. All processing steps were carried out using Matlab (MathWorks Inc, USA), version 2018a. A short script with a set of processing commands can be found in Appendix D.

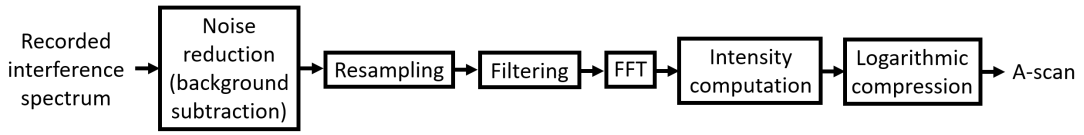

Figure 5.7: Data processing chain in the spectral-domain optical coherence tomography system.

5.4.3 Simulations

Monte Carlo simulations were performed using Matlab (MathWorks Inc, USA, Release 2018a), to evaluate the detected count rate of pixels, depending on parameters such as deadtime, SPAD output pulse width, and number of SPADs per pixel. The simulations allowed to analyse both pulsed and continuous illuminations. Furthermore, SNR values were evaluated based on these count rate simulations. For this, the output of 1024 pixels were simulated for a given optical input with an interferometric spectrum. The simulated spectra were processed using the described data processing methods. The purpose of the simulations is not necessarily to provide accurate estimates of count rates and SNR values, but to reveal the typical behaviour when changing sensor parameters.

In the simulations, the spatial profile of illumination can be set either to be uniform, or Gaussian shaped with variable centre position and width along the pixel height. The power of the beam is adjusted by setting the total number of photons that are gener-

ated for a user defined exposure time. 8 detection elements (SPADs) are considered in a pixel and can be turned on or off, individually. The mean number of photons falling to a certain SPAD is determined by the fraction of the optical beam covered by the SPAD. The number of incident photons on each SPAD is generated as Poisson distributed random numbers with these mean values. After this, an arrival time is generated for each photon. Assuming a constant optical power from a continuous light source, the detection time of a single photon from this source has a uniform distribution (over all possible detection times). Therefore, when simulating a continuous source, arrival times of incident photons are generated with a uniform distribution in time. We can evaluate the number of such uniformly distributed photons, detected within a certain time interval. The detection during the time interval describes a binomial process, where the probability of selection is the ratio of the time interval and the total simulated timespan. The number of photons detected within the time interval follows a binomial distribution in this case. For many generated arrival times, and short time intervals (therefore a low probability for a photon to be measured within the time interval), the number of collected photons can be approximated with Poisson distribution, just as we would expect from shot noise. For simulating pulsed sources, the repetition rate can be varied, but a constant mean number of photons per laser pulse is generated. Photon arrival times are simulated with normal distribution around laser pulses. Dark counts are generated with a uniform distribution, based on real DCR measurements of the red/blue pixels of Ra-II.

After generating all photon arrival times, certain photons are discarded if arriving in the inactive period of a SPAD, which is determined by the user defined deadtime. The inactive period of the SPADs is prolonged with each photon that arrives during this period, according to a paralysable detector model. SPAD output pulses (start and end times of perfect square waves) are calculated for the remaining photons, according to the SPAD pulse width set by the user. Pulses from active SPADs are summed in an OR fashion. FF and finite detection efficiency of the simulated pixel were not taken into consideration. These can be added to the simulation by means of a simple multiplication factor to the expected number of incident photons. Further details of the simulations can be found in Appendix C.

When simulating the effects of deadtime on the achievable count rate, an exposure time of 100 µs and a SPAD pulse width of 300 ps were used. The simulation was set to use continuous illumination, with a Gaussian beam profile of 23 µm FWHM, centred in the middle of the pixel along its height. All 8 SPADs were set to be active. Arbitrarily chosen optical powers were used with respect to the number of simulated photons. The deadtime was scanned from 30 to 50 ns in 0.5 ns steps in consequent simulations with each optical power.

In a separate set of simulations, it was analysed how the number of active SPADs affect the detected count rate. The DCR parameter of the simulations was set according to the measured DCR or red SPADs. First, a uniform illumination model was used, both from a continuous and a pulsed source. Further simulation parameters consisted of an exposure time of 100 µs, a deadtime of 36 ns, and a SPAD pulse width of 300 ps. In the case of the pulsed source, the simulated repetition rate was 20 MHz. The distribution of the generated photon arrival times was centred at the laser pulses, with a standard deviation of 100 ps. The same parameters, but with a Gaussian beam profile were used to generate the output of 1024 pixels. The incident signal was given by simulations of interferometric fringes, caused by a single reflector. The aim of these simulations was to estimate how the SNR changes in a realistic OCT scenario, in function of the number of active SPADs per pixel. First, SPADs in the middle of the pixel were used only, receiving most of the Gaussian beam, then increasingly more and more SPADs were enabled, towards the top and the bottom of the column of SPADs in the pixel.

Lastly, the effect of shortened SPAD pulse width was simulated, both with continuous and pulsed illumination. For this, the same parameters were used as described above, except for the pulse width scanned over a range of 0.2 to 2.6 ns.

5.4.4 Spectrometer alignment

After the simulations, measurements were carried out with Ra-II in the optical setup, using different SPAD configurations and parameters to evaluate their effect on the OCT performance. A crucial step in all the measurements (described below) was to make sure that the optical setup is well aligned. Care was taken to ensure that the sensor is in the focal plane of the focusing lens, and that the sensor and the dispersed optical beam are parallel to each other. For this, spectral lines were recorded with a single line of SPADs enabled at a time. A 512×8 pixel image of the incident beam profile was built after iterating through all 8 lines of SPADs, so that each pixel of the image shows the amount of photons that a pair of SPAD receives (Fig. 5.8a). A Gaussian curve was fitted on the measured beam profile at the centre of the sensor (i.e. pixel 256). The height of the incident beam profile was defined as FWHM of the fit, using the known size of pixels (Fig. 5.8b). The focal point of the spectrometer setup was found where the recorded beam height was minimal (~23 µm).

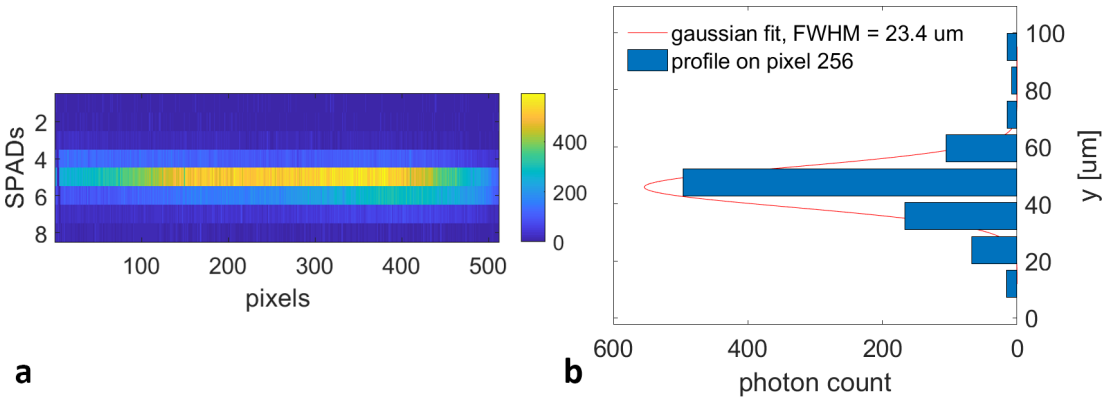

Figure 5.8: Image of the beam profile formed by using each of the 8 lines of SPADs of the Ra-II pixels separately (a). The beam height is defined as the full width at half maximum (FWHM) of a Gaussian fit on pixel 256 (b).

5.4.5 Sensitivity and exposure time

Sensitivity of the OCT system was measured with Ra-II, configured to use 512 pixels, first with all 16 red, then with all 16 blue SPADs enabled. The bias voltage was 26 500 mV for red and 15 050 mV for blue SPADs. In these experiments a continuous illumination was used, provided by the Exalos SLD. A mirror was placed in the sample arm, at a slight path difference with respect to the reference arm mirror. The galvanometer scanner was fixed at the centre position. The optical signals reflected from the sample and reference arm mirrors were attenuated with ND filters (NE40A-B, Thorlabs Ltd, UK).

Several A-scans were recorded while different exposure times were used on Ra-II. The aim of these measurements was to determine the exposure time dependence of the SNR. The exposure time was increased gradually in small steps. With each exposure time 300 lines from the reference arm were acquired and averaged for background subtraction. After this, 300 lines were also acquired with both arms enabled. The lines were transformed to A-scans according to previously described data processing steps. The mirror reflection appears on the A-scans as a single peak. Sensitivity of the system was measured as the height of this peak (signal), compared to the highest point of the base level (noise), plus the attenuation provided by the ND filter of the sample arm. The accurate attenuation of the ND filter was measured prior to the experiment. This was carried out using a power meter, by measuring the power level of the excitation light before and after the filter. The measured SNR values at each exposure time were averaged over the 300 acquired lines to reduce the uncertainty of the measurement.

To compare all three sensors, Ra-II was replaced in the spectrometer setup first with Ra-I and then the PointGrey camera, and sensitivity of the OCT setup was measured again, with an exposure time of 1 ms. With Ra-I, the blue SPADs were used. In the case

of the PointGrey camera, 10 pixels along the height of the pixel array were summed to form spectral lines. First, 2048 pixels of the camera were used along its width, then the sensitivity was measured with 2×2 binning as well. The A-scans were calculated from the spectra, and the sensitivity was calculated from them as described above.

5.4.6 Bias voltage

The effect of increased bias voltage was tested in a similar fashion, using the blue SPADs of Ra-II. No ND filters were installed in any of the interferometer arms, but the interference signal was attenuated in the spectrometer to avoid saturation of the sensor. According to this, the measured SNR describes the intensity of the attenuated mirror reflection, and not the sensitivity of the OCT system. Still, changes in this SNR value reflect how the sensitivity of the whole system varies with the bias voltage. A fixed exposure time of 100 µs and a quench voltage of 3000 mV were used while varying the bias voltage.

5.4.7 Quench voltage

To test the effect of SPAD deadtime on the achievable count rate, the bias voltage of Ra-II was fixed at a level of 15 050 mV (250mV excess bias), and the quench voltage was varied. The measurements were performed several times, each time with a different amount of attenuation in the spectrometer and hence a different optical power falling onto the sensor. The power levels were chosen arbitrarily. In these measurements only the reference arm was enabled.

5.4.8 Number of pixels

The detected spectra of a non-saturating signal were compared when 512 or 1024 pixels of Ra-II are used. A variable ND filter (NDC-50C-4M, Thorlabs Ltd., UK) was placed in the reference arm, which set the optical power falling to the sensor to 100 nW. A bias voltage of 15 050 mV (250 mV excess bias) for blue SPADs, a quench voltage of 3000 mV, and an exposure time of 450 µs were used in this experiment. The sample arm was blocked.

5.4.9 Number of SPADs per pixel

The effect of the number of SPADs per pixel was also tested in different configurations with respect to the number of pixels (512 or 1024), SPAD types (red or blue) and light sources (pulsed and continuous). In each sets of measurements, first, only the SPAD line in the middle of the sensor was used, then more and more lines of SPADs were enabled towards the edges of the pixel columns (see Fig. 5.8). The detected count

rates and SNR values were measured in all configurations. The power of the interference signal falling to the detector was set by the variable ND filter in the spectrometer. The bias voltage was 15 050 mV when the blue SPADs were tested, and 26 500 mV in the case of the red SPADs. SNR values of reflections were calculated with background subtraction and averaging, as discussed above.

5.4.10 Pulse shortener circuits

Tests with varying voltages of the pulse shortener circuit (VBP) were performed with saturating optical powers, using the reference arm only. The light source was the Exalos SLD, but for confirmation of the expected behaviour the measurement was repeated with the Fianium supercontinuum as well.

5.4.11 Pulsed illumination

In order to demonstrate how the repetition rate affects the count rate, the sync output of the Fianium supercontinuum laser was connected to a digital delay generator (DDG) (DDG645, Stanford Research Systems, Inc., USA). One of the output channels of the delay generator was used for triggering time-mask generation on the Ra-II sensor. The rate of the time-mask trigger was altered by the DDG645 unit, by scaling down the rate of the laser sync signal (see Fig. 3.12). With different pre-scale factors, it is possible to set the time-mask trigger rate to any integer fraction of the laser trigger rate. The width of the time-mask was set to cover a single laser period of the supercontinuum laser (50 ns). Masked photon counting was performed on the sensor, using the same exposure time with each different mask trigger rate. Owing to the time-mask, the number of detected optical pulses was ensured to match the number of time-mask trigger pulses during the exposure time.

5.4.12 OCT imaging

OCT images of a fingertip were acquired with all three sensors (Ra-I, Ra-II, PointGrey camera) in the described optical system. The applied light source was the supercontinuum laser in each case. Both with Ra-I and Ra-II, the blue SPADs were used. In case of the PointGrey camera, 10 pixels along the height were summed to from an A-scan. The applied exposure times were 2 ms for Ra-I and 1 ms for Ra-II and the PointGrey camera. The scanning range with Ra-I was 2.4 mm and 2.6 mm with Ra-II, both with 300 scanning points. The frame rate of the PointGrey camera is limited to 1 frame per 10 ms. With this camera frame rate, the OCT B-scan frame rate of large images (many scanning points) would be rather low. To overcome this problem the scan range was reduced to 1.6 mm (200 scanning points). Each A-scan was background subtracted, as discussed before.

5.5 Results

5.5.1 Sensitivity and exposure time

Figure 5.9 shows the measured OCT sensitivity with the red SPADs of Ra-II, linearly increasing with the exposure time (blue curve), in agreement with the predicted behaviour (see equation 5.3). The achieved sensitivity values are plotted on a logarithmic scale as well, which is most often used in OCT (red curve). Sensitivity with 1 ms long exposure times was measured to be 86 dB with the red and 87 dB with the blue SPADs.

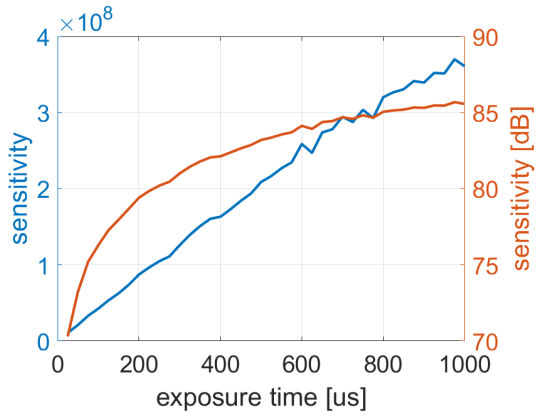

Figure 5.9: Exposure time dependence of sensitivity, plotted both on a linear scale (blue curve) and in dB units (red curve). Sensitivity increases linearly with exposure time (logarithmically in dB units), offering a trade-off between sensitivity and frame rate.

The measured sensitivity with the blue SPADs of Ra-I was slightly above 80 dB. The sensitivity with the PointGrey camera is 15 dB higher than with Ra-II without binning, and 20 dB higher with 2×2 binning.

5.5.2 Bias Voltage

Figure 5.10a depicts the total number of detected counts from the reference arm versus the bias voltage, measured with Ra-II. The increase in the number of counts is evident as the bias voltage increases. The measured SNR of the interferometric spectrum (both arms enabled) in function of the bias voltage is, however, almost constant above the breakdown voltage (Fig. 5.10b). A closer look reveals an ever slight SNR increase at the beginning, and a minor decrease above a certain voltage (inset in Fig. 5.10b). A negligible SNR increase was observed with the red SPADs as well with no decline (not shown). The lack of the minor decline is considered to be caused by the lower DCR of red SPADs. When using the pulsed source, the increase of counts with bias voltage was less intense, and the decline of SNR was slightly stronger with both SPADs (not shown). The reason for this is that with a pulsed source, there is

an upper detection limit of one photon per laser pulse. These results suggest, that even though increasing count rates should be accompanied with increasing sensitivity at higher bias voltages, the increase in DCR prevents benefiting from this, especially when the count rate increase is minor compared to DCR.

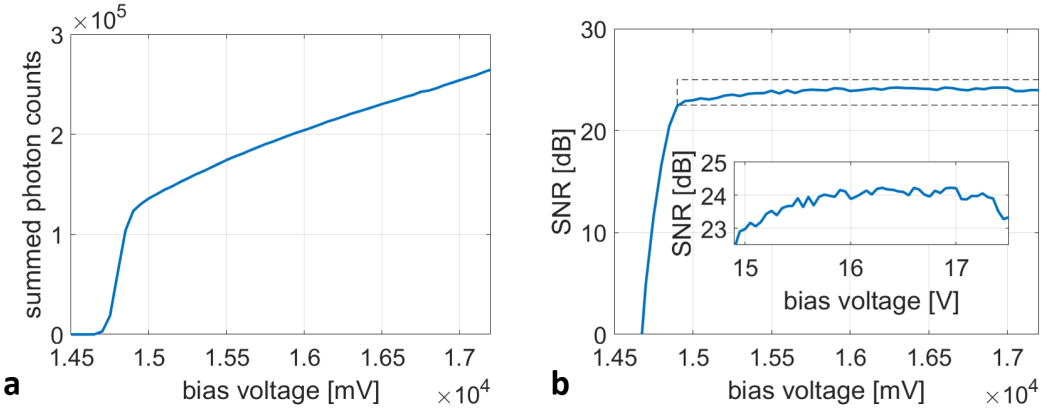

Figure 5.10: The total count rate detected with blue SPADs of Ra-II starts increasing linearly with the bias voltage, above the breakdown voltage (~14 800 mV) (a). The measured signal-to-noise ratio (SNR) cannot benefit from the higher count rate, due to increasing dark count rate (b). The inset figure suggests that the DCR increase is even higher than the increase of detected counts.

5.5.3 Quench voltage

According to the simulations, when the incident count rate is high (saturating), then the detected count rate increases with shorter and shorter deadtime (yellow and purple curves in Fig. 5.11a). Varying the deadtime has no effect if the count rate is sufficiently low (with respect to the saturation level) (blue and red curves in Fig. 5.11a).

Measurements with different quench voltages show similar results. With low incident count rates (when the detection is not deadtime limited), changing the quench voltage does not affect the number of detected photons (blue and red curves, Fig. 5.11b). In the case of saturating optical powers, an increased quench voltage increases the number of detected counts, and the level of saturation (yellow and purple curves in Fig. 5.11b). This increase stops when the corresponding deadtime reaches the reciprocal of the saturating count rate, or when the deadtime cannot be shortened further by increasing the quench voltage (due to limitations of the applied electronic circuit). Note, that one of the reasons why the shape of the simulated and measured curves are different is because the simulated results are plotted in function of deadtime, and the measurement results are plotted in function of the quench voltage.

Increasing the quench voltage has no effect with pulsed light sources at a repetition rate lower than the saturation count rate (i.e shorter deadtime than laser pulse period).

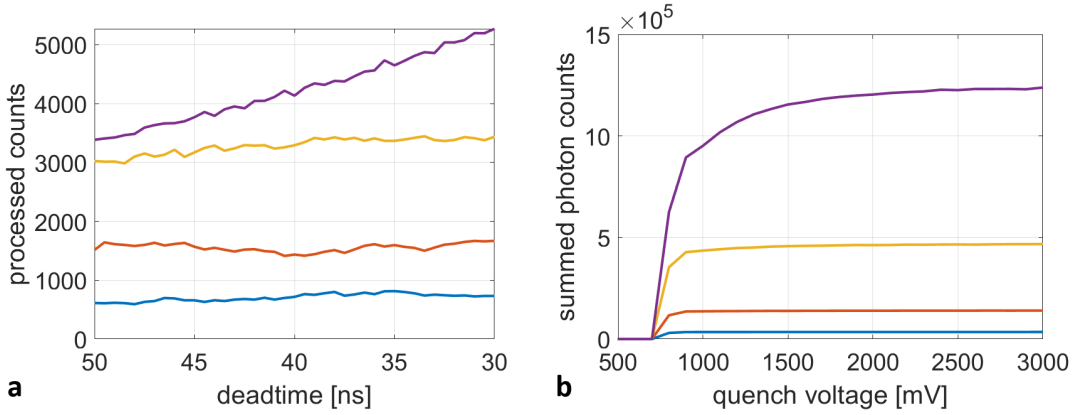

Figure 5.11: Detected counts on a single pixel in 100 µs, based on Monte Carlo simulations of photon arrivals (a). The simulations involved different scenarios with respect to the deadtime of SPADs, and the number of incident photons (optical power). Measurement of the detected number of counts on the entire sensor in function of the quench voltage, with several different optical powers applied (b). Note, that the counts are expressed in function of different parameters in the two plots (deadtime in the simulations and quench voltage in the experiments).

5.5.4 Number of pixels

The measured spectrum from the reference arm when the incident power is set to 100 nW is plotted in Fig. 5.12, when 512 or 1024 pixels are used. For direct comparison, the two curves are plotted on top of each other, using different x-axes. The plot confirms that the number of detected counts are approximately halved when twice the number of pixels are used. In a separate measurement with strong optical powers, it was also confirmed that doubling the number of pixels allows twice the optical power to be processed without saturation.

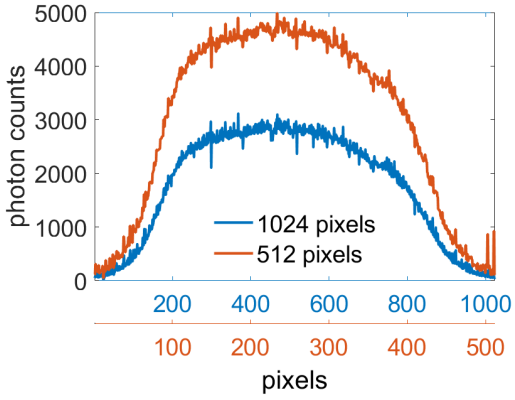

Figure 5.12: Detection of an optical signal with 512 or 1024 pixels of Ra-II.

5.5.5 Number of SPADs per pixel

The simulated count rate in function of the number of SPADs per pixel is shown in Fig. 5.13a for a continuous, and Fig. 5.13b for a pulsed source. In both cases, the simulated pixel was uniformly illuminated, and its DCR was set according to the real DCR of red SPADs. In both cases, the detected count rate starts increasing linearly with the incident count rate. With a continuous source (Fig. 5.13a), the curves depict the typical behaviour of SPADs. The SPAD pulses with finite width start to overlap at a high number of incident photons. Photons with overlapping pulses are detected as one count. The more frequently this happens (increasing with the incident photon flux), the further the transfer curve will be from a linear response. Eventually, the incident count rate becomes so high, that multiple photons hit a single SPAD during its deadtime. At these circumstances the SPADs cannot be reinstated to the active state, preventing detection and counting of the photons.

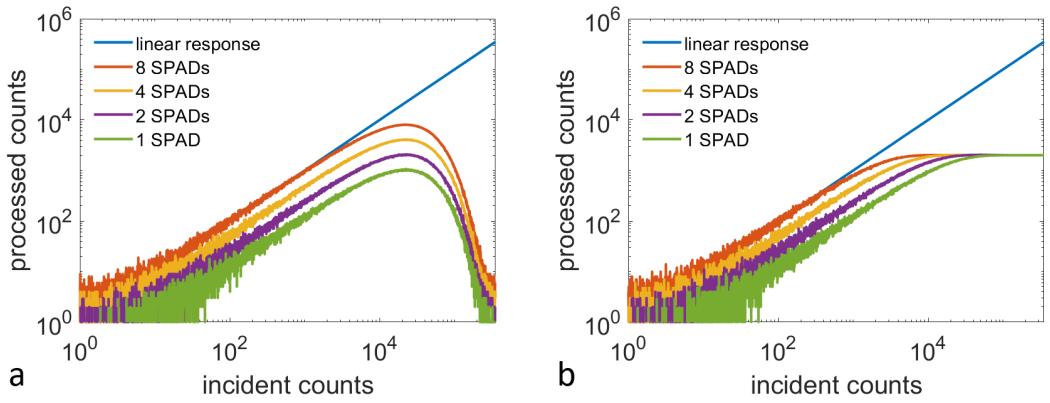

Figure 5.13: Detected counts versus incident counts of a pixel based on simulations for a continuous (a) and a pulsed (b) light source. The curves show a typical behaviour of paralysable detectors with a continuous source (a). When a pulsed source is used (b) the count rate is limited by the repetition rate, allowing a maximum of one photon detection per laser pulse, independently of the number of detection elements.

The case with a pulsed source (Fig. 5.13b) is slightly different. Here, the detected count rate increases, as long as the expected number of photons per laser pulse is less than one, in average. Statistically, the number of photons detected throughout many laser pulses (i.e. over the exposure time) reflects the strength of the incident signal. As the incident power increases, after a certain point the average number of photons per laser pulse will be more than one, and so multiple photons will hit the pixel from a single laser pulse. Since the photons of the sharp light pulse reach the pixel roughly at the same time, their respective SPAD pulses overlap. As a result, the maximum number of detected photons is equal to the number of laser pulses in the exposure time. The simulations also show how the number of detected counts for a

given optical input increases with more and more SPADs enabled. On a linear scale, this would be represented in different rise of the curves (that is, different sensitivity of the pixel). On a log-log scale, this is revealed by the offset between the curves. In Fig. 5.13, the transfer curves belonging to exponentially increasing SPAD numbers are equidistantly spaced on the log scale. This suggests, that on a linear scale the increase in the rise of the curves would be linear with the number of SPADs, therefore the detector sensitivity linearly increases with the number of SPADs per pixel (at low incident count rates).

Simulated and measured SNR values in function of the number of SPADs per pixel are plotted in Fig. 5.14a and b, respectively. The beam profile in the simulations was set to be Gaussian shaped, similar to that determined from real measurements. Both simulations and measurements were performed with all four combinations of blue/red pixels, and continuous/pulsed sources. Red/blue pixels in simulation context means that the simulated DCR was based on the real DCR of either the red or the blue SPADs of Ra-II. The plotted results are limited to the combinations of blue pixels with a pulsed source, and red pixels with a continuous one. The behaviour of the two types of SPADs is seemingly different, however no significant difference was observed between the light sources when used with the same type of SPADs, neither in simulations, nor in measurements. SNR with blue SPADs gets worse after an initial rise, as the number of detection elements per pixel increases. The initial rise comes from an increased incident count rate, which has a diminishing effect as the tail of the beam profile vanishes (refer to Fig. 5.3b). The DCR noise keeps increasing, and adds –3 dB to the SNR with each doubling of SPADs. In the case of the red SPADs, the DCR is low enough not to cause a significant deterioration, and the SNR flattens. Simulations (Fig. 5.14a) and measurements (Fig. 5.14b) both reveal the same trend with the two types of SPADs, even though the actual SNR values are different between simulation and measurements.

The measurements reveal a slightly different behaviour when the sensor is configured to use 512 or 1024 pixels. Independently of the types of SPADs and the light sources, the achieved SNR is slightly higher in 1024 pixel mode, when only a few SPADs per pixel are used. However, as there are more and more SPADs enabled, the SNR of the 1024 pixel mode drops below that of the 512 mode.

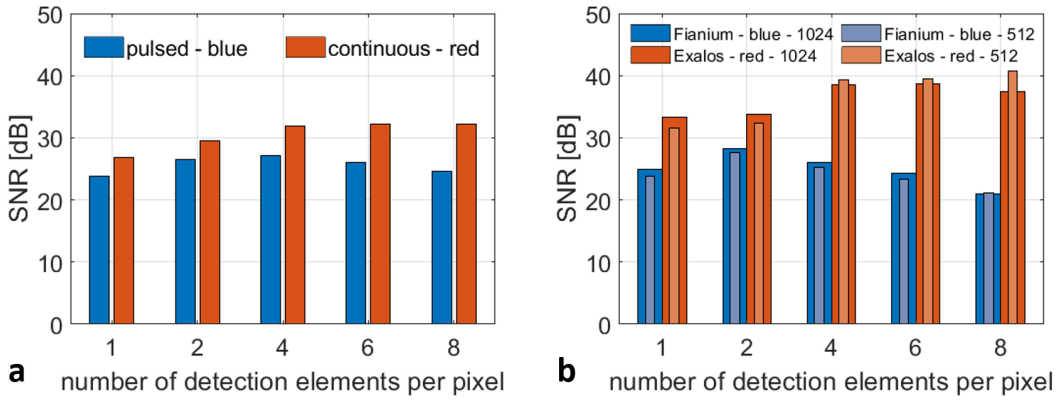

Figure 5.14: Simulated signal-to-noise ratio (SNR) in function of the number of detection elements per pixel, using pulsed illumination and the blue SPADs, or continuous illumination and the red SPADs (a). Measured SNR in similar scenarios, both when a 512 pixel or a 1024 pixel configuration is used (b). The standard deviation of the simulated SNR was less than 0.5 dB in each case, whereas it was less than 1 dB in the case of the measurements. Note, that a detection element means two SPADs in 512 pixel mode, and a single SPAD in 1024 pixel mode.

5.5.6 Pulse shortener circuits

Simulated results of shortening the SPAD pulses and measurements with increasing the monostable voltages (VBP) are shown in Fig. 5.15a and b, respectively. The simulated results are plotted against the reciprocal of the pulse width, so that simulation and measurement results can be compared. In both cases the photon counts are scaled to what is achieved with the longest pulses (highest VBP voltages). The results show, that with saturating optical signals, shortening the SPAD pulses increases the number of detected photons and the saturation level. Above a certain VBP voltage (~0.9 V) the monostable circuit does not affect the pulses. Below this value, pulses gets shorter with decreasing VBP. The count rate increase stops when the pulses are short enough so all absorbed photons can be resolved, or at the limits of the pulse shortening circuit.

5.5.7 Pulsed illumination

The results in Fig. 5.16 confirms that higher repetition rates of the source infer higher count rates, which indicates a higher sensitivity that can be achieved. Indeed, one could think of continuous illumination as pulsed illumination with infinite repetition rate. The upper limit of count rates in either case is imposed by the deadtime of the SPADs.

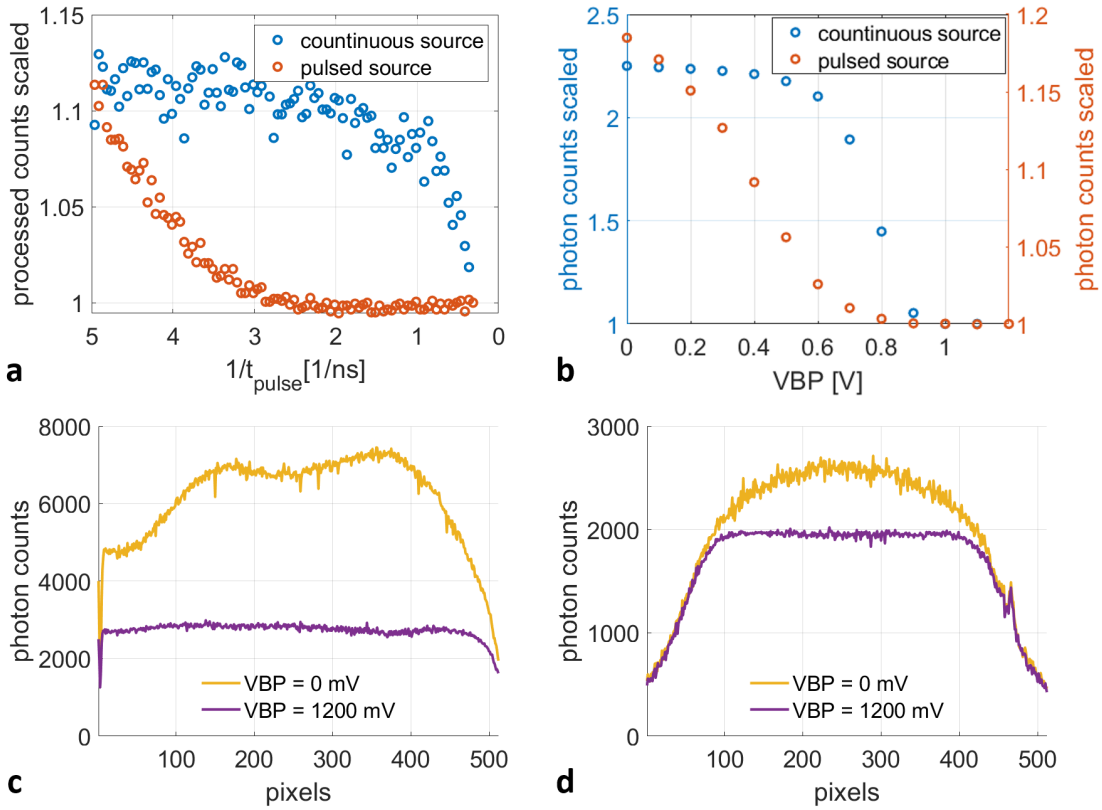

Figure 5.15: Simulated (a) and measured (b) number of counts in function of the SPAD pulse width, with continuous and pulsed illumination (a,b). Mismatch in the increase of detected counts between simulation and measurement is most probably caused by inaccurate simulation parameters. When the SPAD pulses are wide, a high power of illumination (either coming from a continuous (c) or pulsed source (d)) saturates the detector, and the detected counts on the pixel cannot increase above a maximum number (purple lines). The same power of illumination can be detected with short SPAD pulses (yellow lines), showing that the saturation level increases in this case.

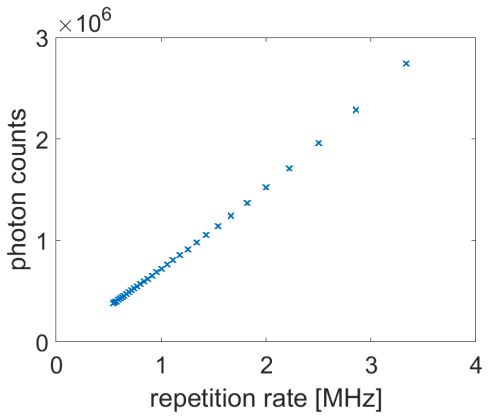

Figure 5.16: The number of detected counts as a function of the repetition rate of the pulsed light source.

5.5.8 OCT imaging

OCT images of a fingertip taken with the three sensors are depicted in Fig. 5.17 (Ra-I, Ra-II and the PointGrey camera on a, b and c, respectively), scaled to matching sizes of the field of view. The first obvious difference between the three images is the depth of field, which is related to the number of pixels in each case (Ra-I - 256 pixels, Ra-II - 1024 pixels, PointGrey - 2048 pixels). Images with Ra-II and the PointGrey camera both show a large depth with no features, since the back-scattered light from these depths is too weak and dispersed for detection (to put it in another way, the correlation between light from this depth and the source is below the noise limit). Nonetheless, the entire image depth is presented here, to demonstrate how the depth depends on the sampling of the spectrum (i.e. the number of pixels). Adjusting the depth region of interest over the pixels of the B-scan can be achieved by changing the width of the detected spectrum (and therefore the axial sampling interval, see equation 2.16).

Vertical lines in the Ra-II image are caused by occasional reading of erroneous lines from the sensor, caused by a bug in the Opal Kelly software library. It should also be pointed out, that while having a fingertip as a test target is convenient, for the comparison of different configurations a steady, standard test target would be more appropriate.

5.6 Discussion and conclusions

5.6.1 Sensitivity

The measured sensitivity of Ra-II (87 dB) is slightly higher than that of Ra-I (80 dB), both measured with an exposure time of 1 ms. The most probable reason for this is the higher number of SPADs per pixels, and improved SPAD design. Even so, the sensitivity with Ra-II is reasonably lower than that with the PointGrey camera. The quantum efficiency (probability of turning photons to electrons) is similar with CCD, CMOS photodiode and CMOS SPAD technologies, however, SPAD sensors have a low FF, and consequently low photon detection efficiency (PDE). The sensitivity differences are noticeable in the OCT images as well. The image with Ra-I is inferior to the other two, even with twice the exposure time. Ra-II yields a much better quality image. Still, the most details along the depth can be seen with the PointGrey camera.

Sensitivity can be increased through longer acquisition of the information embedded in the interference spectrum. However, long exposure times can be disadvantageous due to fringe washout. No such deterioration of the signal was observed using 1 ms exposure time with the Ra-II sensor. Still, such acquisition time provides a significantly lower line rate than that of recent spectral-domain OCT systems with up to million

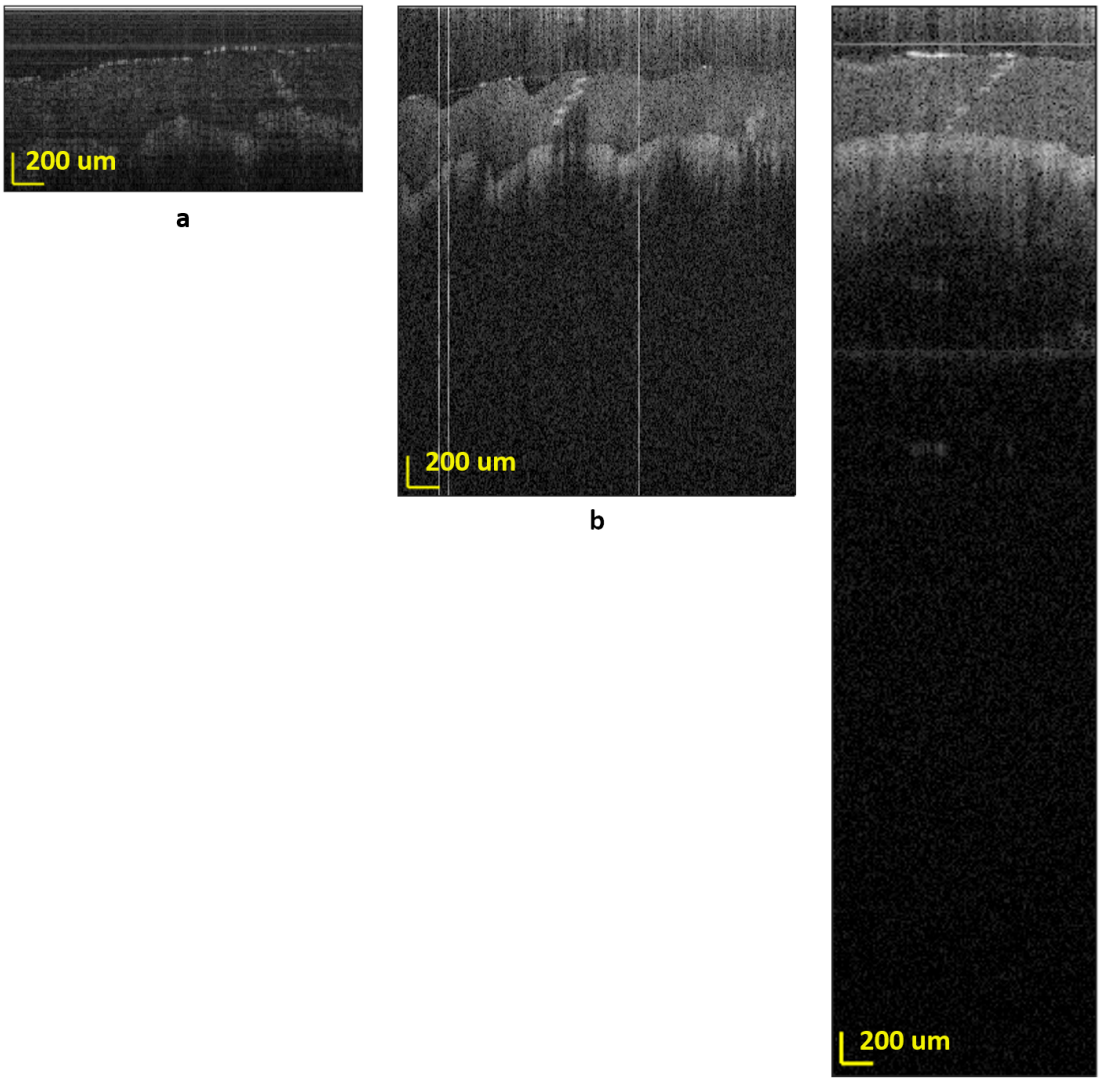

Figure 5.17: Optical coherence tomography images of a fingertip taken with different sensors. Ra-I, using the blue SPADs, and 2 ms exposure time per line (a). Ra-II, using 1024 pixels and blue SPADs with an exposure time of 1 ms (b). The Point Grey camera with 1 ms exposure time (c). The light source was the supercontinuum laser in each case.

frames per second (Barrick *et al.*, 2016). High frame rate is key in video-rate OCT imaging, particularly with three dimensional volumetric images. In order to maintain good signal at short exposures, the detector needs to have high sensitivity of photon detection at the source wavelength.

SPAD sensor parameters were analysed using the Ra-II sensor to achieve the best possible sensitivity. While increased count rates are achieved in many cases, these are only beneficial if not superseded by a similarly increased noise. Red SPADs were designed to be more efficient for the detection of NIR photons. Even though having a

higher photon detection probability (PDP), the active area of the red SPADs on the Ra-II sensor is about 2.6 times smaller than that of the blue SPADs, leading to similar PDE of the two types of pixels (see Fig. 3.23). Red SPADs are also superior with respect to the DCR noise they yield. Nevertheless, the measured sensitivity was similar with the two types of SPADs (86 dB and 87 dB for red and blue, respectively). No qualitative difference could be seen in OCT images either, taken with the two types of SPADs. These can be found in Appendix E, where a number of OCT images are collected, acquired in different configurations and scenarios with respect to the sensor and the light sources.

Ra-II allows different pixel configurations, either having 512 pixels of 2×8 SPADs, or 1024 pixels with 8 SPADs in a column. The different modes affect the spectral sampling, and consequently the axial range of the OCT images, as discussed in section 5.3.4. This is clearly demonstrated in the OCT images in Fig. 5.17, where the same spectrum is detected with devices having more and more pixels. It was also shown, that the number of detected counts are also different when detecting a certain optical power with 512 or 1024 pixels (Fig. 5.12). When considering the expected SNR in the two cases, the detected count rate and DCR both need to be considered. The assumption is that in 512 pixel mode, the detected number of counts is two times higher, and therefore signal to shot noise ratio is higher as well. DCR noise in 512 pixel mode also doubles, which cancels out the SNR improvement. The measured SNR in 512 or 1024 pixel modes are, however, slightly different. This may be explained by the following. Let us assume a certain optical input from a continuous source, falling on a column of 8 SPADs. If the optical signal is not too strong, the majority of the incident counts can be detected by the SPADs, without missing too many photon arrivals because of arriving too close to each other. If we now consider that two columns of 8 SPADs are joint (this is the 512 pixel mode), the OR gate where the SPADs are combined need to process photons that arrive two times more frequently (in average). Consequently, the number of missed photons (due to arriving too close to each other) increases. Therefore, depending on the incident optical intensity as well, the 1024 pixel mode detects a slightly higher number of photons, and achieve a slightly higher SNR.

The effect of the number of SPADs per pixel on the SNR was studied through simulations and measurements. The trends observed in simulated and measured results are similar, even though the simulated SNR values are consistently lower than the measured ones (Fig. 5.14). The most probable reason for this is inaccurate parameters applied in the simulations. A notable difference is that no initial SNR increase was observed in the measurements with red SPADs. This is believed to be caused by the two different types of red SPADs, and that the level of bias voltage was not sufficient to exceed the breakdown voltage of both types. The general observation from the

results is that it is advantageous to deploy more and more SPADs for higher SNR, provided that they all receive a reasonable fraction of the optical signal (that is, their contribution to the signal power is higher than it is to noise variance). As an example, using two SPADs instead of one increases the SNR, but enabling too many high DCR blue SPADs (at the edges of the pixel, with minimal detected signal from a Gaussian beam) decreases the SNR, as shown in Fig. 5.14. Influenced by this finding, further experiments were carried out with the continuous light source, where a cylindrical lens was placed in front of the sensor. The lens was used to stretch the focused beam along the height of the pixel, providing a more relaxed count rate on the middle of the pixels and covering more SPADs by the optical signal. According to the results (not shown), a slight increase in the count rate and the sensitivity of the system is achievable. The drawback, however, is that even though the cylindrical lens is expected to modify the beam only along one axis (along the pixel height), the photons are somewhat diffused along the other optical axis as well (i.e. along the spectrum), limiting the spectral resolution. This leads to a lower visibility of high frequency fringe modulation, and a significant drop of sensitivity for signals from deeper regions of the sample. For this reason, the method for stretching the optical beam to cover more SPADs of the pixel was not investigated further.

As mentioned, a sensor configuration with 1024 pixels is expected to have a slightly higher SNR than that with only 512 pixels. This is indeed the case, when only a few number of SPADs per pixel are enabled (see the difference between narrow and wide bars in Fig. 5.14b). However, as more and more lines of SPADs are activated, the SNR in the 1024 pixel mode drops below that of the 512 pixel mode (six and eight detection elements in Fig. 5.14b). This behaviour was observed in the measurements with both types of SPADs and light sources. It is believed, that the phenomenon is caused by DCR variation over the pixels. The expected value of per pixel DCR is more uniform in the 512 pixel mode, where DCR of the pixel is 'averaged' over a higher number of SPADs. Presumably, per pixel DCR variation adds broadband noise to the respective A-scans. In 1024 pixel mode, this additional noise is somewhat higher and slowly wins over the benefit of increased signal to shot noise ratio. This theory requires further analysis, e.g. by measuring the power spectral density (PSD) or the serial correlation of DCR over the pixels. The variance of DCR across the pixels in 512 and 1024 modes could also imply if either mode is detrimental in this regard.

5.6.2 Saturation and harmonics

The saturation level of a SPAD is decided by its deadtime. In the case of pixels with multiple SPADs, the limiting factor becomes the SPAD pulse width (Gnecchi *et al.*, 2016). This behaviour is independent from the exposure time. With all blue SPADs enabled, the optical power from the Exalos source was about 1.5 µW when Ra-II saturated.

The signal saturation in SPADs is clipping the interferometric fringes detected which, in turn, introduces harmonics in the Fourier transformed data (Raele *et al.*, 2014). It is not just clipping that is problematic, but strong optical signals close to saturation as well. Close to the saturation level of a single SPAD, several incident photons are missed, due to increasingly more time spent in an inactive state of the device. This introduces a deviation from a linear response, which increases with the incident count rate. Simulations confirmed, that when a continuous source is used, a multiplexed pixel with several SPADs increases the level of saturation. This can be seen in Fig. 5.13a, where the maximum of the transfer curves increases with the number of SPADs. This is only true, if the same optical power is dispersed over more and more pixels. In contrast, when forming 512 or 1024 pixels to receive the same spectrum, there are more photons missed due to counting loss with 512 pixels (which detect twice the optical power than pixels in the 1024 mode), even though they have twice the number of SPADs, as discussed already. In this regard, lower optical powers are preferred. It should be emphasised, that this phenomenon is a property of SPAD devices. In general, OCT requires strong illumination for shot noise limited operation. To benefit from strong incident light, SPAD detectors need to be able to process many photons with linear response. This is especially true, when the illumination is not uniform. In these cases, even if the optical power is tolerable in general (which we can see from the fact that the detected count rate increases with further increasing of the optical power), SPADs of the pixel receiving most of the incident light start to saturate, and therefore the increase of the detected counts in not linear with the power increase. A non-linear SPAD and pixel response introduce the same harmonics in the transformed data as when clipping the fringes. During the experiments, harmonics started to be noticeable when the incident count rate was above than 10 percent of the saturation count rate. Harmonics were often observed with the Ra-I sensor, which had a low saturation level due to the lower number of SPADs per pixel. Repeated structures in the OCT images due to sensor saturation are depicted in Fig. E.2 of Appendix E.

Simulations and measurements showed that shorter SPAD pulse widths increase the count rate. This behaviour is predictable with a continuous source (Fig. 5.15c). When the pulsed source is used, SPADs of a pixel are expected to fire at the same time. Hence, no increase is expected in the detected counts when changing VBP. However,

a slight increase of the detected counts is observed with the pulsed source as well (Fig. 5.15d). This could happen if the laser pulse duration is longer than the generated SPAD outputs. While simulation and measurement results are seemingly similar, it has to be pointed out that the actual measurements showed a higher increase of detected counts, especially in the case of continuous illumination. This is most likely due to inaccurate simulation parameters.

No hard clipping is observed with the red pixels, even for strong, saturating signals. This is believed to be caused by crosstalk between red SPADs (shown in section 3.4.5) and potentially increased afterpulsing. For example, the detected count rates with a pulsed source and red SPADs can increase well above the laser repetition rate (even though the maximum count rate should be equal to the laser repetition rate). The number of extra counts (not directly caused by incoming photons but from previous and neighbouring avalanche breakdowns) increases with the optical power. This effect could be beneficial from a spectral interferometry point of view, since it increases the modulation depth of the interferometric fringes, just as if the detector had internal amplification. However, the extra counts from crosstalk and afterpulsing do not scale linearly with the incident optical power, and hence they introduce harmonics in the Fourier domain, and cause repeated depth profiles, accordingly. Another likely drawback is the loss of spectral resolution. Nonetheless, appropriate evaluation of the idea requires further considerations and thorough testing.

In the case of CCD and CMOS cameras, the saturation depends on the well depth. Essentially, this sets an upper limit on the integral of the detected signal. As a result, short exposure times can tolerate stronger optical signals, but long exposures might saturate the device even at lower optical powers. The PointGrey camera, which was tested against the Ra sensors, saturates when the optical power is about 2 to 2.5 µW from the Exalos light source, at 50 µs exposure time. At an exposure time of 1 ms, this value is only around 100 nW. When 2×2 pixel binning is applied for increased sensitivity, this value drops to its quarter. Therefore, it is beneficial for strong optical signals to use short exposure times. On the other hand, when the exposure time is short, faint optical signals (e.g. backscattering from deep regions of tissue) integrate a low charge only, limiting the sensitivity of the system. This is the same type of trade-off as the one discussed with the sensitivity versus saturation level of SPAD sensors. With respect to harmonics, the major limiting factor with the PointGrey camera was not a non-linear pixel response, but the fact that spectral lines were formed of several lines of the pixel array summed. The detected counts are non-linear to the incident counts when a single line of pixels in the middle of the optical beam is in saturation.

To tackle strong optical signals without non-linearity and saturation artefacts, SPAD based detectors need to have a short deadtime. While saturation levels are obviously

worse for SPAD sensors compared to more common photodiode structures, it is very likely that future SPAD devices will be more integrated and will approach or improve on saturation levels.

5.6.3 Continuous and pulsed illumination

Time-resolved detection usually requires pulsed illumination. Modulation is also used in frequency domain fluorescence lifetime imaging microscopy (FLIM), but this is less accurate than TCSPC. One of the drawbacks of pulsed illumination is that photon emission is confined in a short time interval, therefore the light pulses have high peak powers. This may be problematic in medical applications where the applicable average excitation power is often limited by safety considerations.

For CCDs and CMOS cameras the temporal distribution of photon arrivals is not relevant. For SPAD based pixels, however, pulsed sources introduce limitations. Even when several SPADs are used to form a pixel, the pixel cannot detect more than one photon of each short laser pulse. Consequently, the highest number of detected counts is defined by the number of laser pulses during the exposure time. The peak count rate (and the saturation level), therefore, is limited by the light source repetition rate. Since the peak count rate is lower than that with a continuous source, the OCT system is more prone to saturation and harmonics when a pulsed source is used and SPAD detectors.

It was shown, that the count rate can be increased slightly above the source repetition rate if the SPAD pulses are sufficiently compressed. Significant count rate increase can only be achieved by increasing the repetition rate of the pulsed source.

5.6.4 Conclusion

While recent advancements of SPAD sensors are enabling new applications, several properties of SPAD sensors still need improvement. The main limitations are sensitivity in the NIR regime and the ability to resolve strong optical signals. There are no fundamental stumbling blocks in having low noise and efficient CMOS SPAD sensors. Bursts of photons could be detected by adding more SPADs per pixel with shorter pulse widths. Multiple SPADs in a column but with individual counters could enable the detection of multiple photons per laser pulse. Sensitivity can be tackled by stacking them in 3D. No prior work has focused on time-resolved single-photon detection in OCT and the aim throughout this chapter has been to showcase the strengths CMOS SPAD sensors in interferometry. The next chapter discusses the combination of temporally and spectrally resolved detection with SPAD sensors, focusing on novel applications in OCT and low-coherence interferometry in general.

Chapter 6

Time-resolved Spectral-Domain Optical Coherence Tomography

6.1 Introduction

There has been a growing interest in detectors with high sensitivity and low noise, especially in optical applications where photon budgets are low. Complementary metal-oxide-semiconductor (CMOS) single-photon avalanche diode (SPAD) based sensors have been successfully deployed to face these challenges in an ever growing range of applications, especially where strict timing of photons is required. In optical coherence tomography (OCT) systems the detected optical signals from backreflection and backscattering are typically strong, not requiring single-photon resolving capability. The traditional detectors in these systems have been charge-coupled device (CCD) and CMOS based sensors. Noise of the sensor negatively affect the OCT performance. Shot noise limited operation in OCT has been achieved with balanced detection techniques. Single-photon counting was introduced to OCT by deploying superconducting single-photon detectors (SSPDs) and SPADs (Mohan *et al.*, 2008). The avalanche effect in single-photon counting (SPC) removes the readout noise found in all photodiode based sensors with an analogue-to-digital converter (ADC).

In the previous chapter the applicability of SPAD based line sensors was investigated in spectral-domain (SD) OCT. The measured sensitivity of the OCT system with Ra-II was 87 dB with 1 ms exposure time, which is 15 dB lower than that with a commercial CMOS imager, having a quantum efficiency (QE) of 30 % at 850 nm. Despite the capability of single-photon counting, SPADs need a higher detection efficiency in the near infrared (NIR) region (see Fig. 3.23). Handling strong optical signals and bursts of photons is another challenge for SPAD sensors in OCT. The main advantage of SPAD based sensors is the unprecedented capability of timestamping photons with high resolution on hundreds of parallel channels. For each detected photon, time-correlation can be established with respect to laser pulses or other events, opening new avenues in studying the backscattered spectra. Therefore, there is untapped

potential in combining low coherence interferometry with arrays of time-resolved single photon counters for advanced spectroscopy applications.

6.1.1 Time-resolved spectral-domain optical coherence tomography

This chapter investigates the idea of time-resolved detection of interferometric spectra, focusing on potential applications in SD OCT. Figure 6.1 depicts the concept of the combined detection scheme. Time-resolved detection at the current state of SPAD technology is capable of resolving distances of a few millimetres, using pulsed light sources and time-of-flight measurements of the reflected pulses (Niclass *et al.*, 2005; Burri *et al.*, 2016). A time-resolution of femtoseconds, in theory, could reveal micrometre-resolved depths. Currently, differentiation of distances with micrometre resolution requires interferometric techniques. While providing high depth resolution, low-coherence interferometry is limited to a total depth of few millimetres. The concept of our proposed approach is to combine time-resolved detection and spectral low-coherence interferometry, using CMOS SPAD based sensors in interferometer based optical arrangements. With a pulsed source, the sensor assigns the interferometric spectrum of each pulse to coarse depth regions of a few millimetres, based on its time of arrival.

As an example in Fig. 6.1, a time resolution of 20 ps would provide 3 mm resolution along the depth (assuming that the optical signal travels with the speed of light in vacuum). The interferometric spectra are to be used to fine-resolve the coarse depth regions through the inverse Fourier transform, providing micrometre resolution within each coarse depth region. In order to construct interferometric spectra at each time-bin, multiple reflective surfaces are needed along the reference arm. The zero depth of the fine resolved depth profile in each bin is set out by the reflective surface belonging to the bin. For this, closely spaced semi-reflective surfaces are to be used (Fig. 6.2).

The proposed technique has potential benefits across multiple applications and is intended to open up new ways in studying optical phenomena. As an example, SD OCT systems have a limited imaging depth, due to the Hermitian symmetry property of the Fourier transform and the finite spectral resolution of the sensor. The interference spectrum is real-valued, hence its inverse Fourier transform is redundant. Several studies have focused on improving the imaging depth. One way to double the imaging depth is to reconstruct the complex representation of the inverse Fourier transform (full-range complex (FRC) SD OCT). Subsequent phase-shifted measurements of the interferogram could be performed for this, at the expense of the line-rate and the imaging speed. Phase shifts can be produced by changing the reference arm path length. However, systems deploying electro- and acousto-optic modulators and piezo transducers as phase-shifters (Fercher *et al.*, 1998; Wojtkowski *et al.*, 2002)

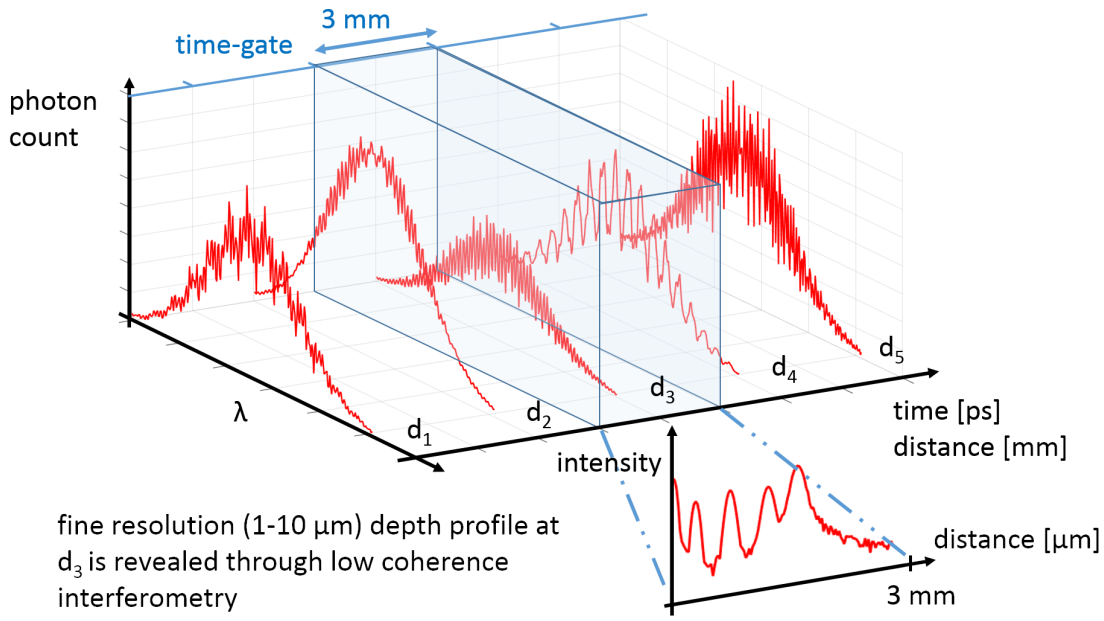

Figure 6.1: Combining time-resolved detection and low coherence interferometry. Time-resolved detection with ps resolution assigns the detected spectra to depth regions with mm resolution. Each spectrum along the depth can be resolved then to a fine depth profile with µm resolution.

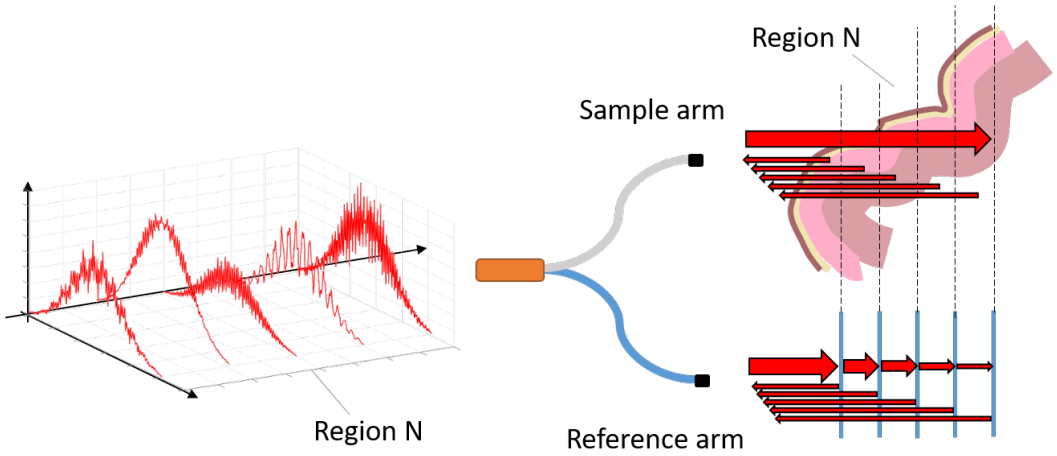

Figure 6.2: The construction of interferometric fringes at each time bin can be achieved, conceptually, by tightly spaced semi-reflective surfaces in the reference arm. The position of each reflector assigns the zero path difference for each interferometric spectrum detected in the time bins, and therefore needs to match the depth belonging to the time bin resolution. Each reference arm reflection interferes with the sample arm signal from the same depth, forming the spectrum revealing the fine resolved depth profile in the respective depth region.

have increased complexity and cost. A more effective phase shift was proposed by introducing a slight displacement between the probe beam and the pivot axis of the galvanometer scanner mirror which is used for transversal scanning (Götzinger *et al.*, 2005; Leitgeb *et al.*, 2007; An and Wang, 2007). The doubled depth range of FRC OCT systems (Götzinger *et al.*, 2005; Leitgeb *et al.*, 2007; An and Wang, 2007; Baumann *et al.*, 2007; Jungwirth *et al.*, 2009; An *et al.*, 2010) is still not satisfactory for many applications. Imaging of even larger depths was achieved with multiple channel configurations (Grajciar *et al.*, 2008; Zhou *et al.*, 2009). The drawbacks of these systems include an increased size, cost and complexity of the optical arrangements. In our proposed method the imaging depth covered by the sensor can be multiplied by the number of time bins. With a given bin resolution, the number of bins and therefore the number of coarse depth regions and the total imaging depth is governed by the repetition rate of the pulsed light source. The penetration of light into tissue is still limited by scattering, determining how well the structures below the top surface can be seen. The depth that is accessible for the sensor, however, is increased and not limited by sampling of the spectrum. Therefore, a certain thickness of the sample can be studied (limited by scattering), independently of the position of the top surface along the total depth seen by the time-resolved sensor (Fig. 6.2).

6.1.2 Time-gated spectral-domain optical coherence tomography

The finite dynamic range of the detector is often problematic in SD OCT. Limited dynamic range of the sensor reduces the visibility of weak reflections in the presence of strong scattering interfaces along the depth. This is even more of a problem when the strong backreflected signal saturates the detector (LaRocca *et al.*, 2011). Different solutions have been proposed to overcome this problem based on post processing the acquired data (Huang and Kang, 2012; Lee *et al.*, 2014), automatic control of the reference signal intensity (Kim *et al.*, 2015), or using a dual-line CCD camera (Wu *et al.*, 2014). Time-gating unwanted reflections in SD OCT is one example where the combination of time-resolved detection and interferometry can be advantageous. Time-resolved detection allows differentiation of multiple reflections along the optical path, and the removal of the signal of unwanted reflections. In these cases, when a single depth region of interest is to be recorded only, time-gated photon counting can be applied instead of recording full time-resolved spectral histograms. In the previously discussed explanation, time-gating of SD OCT signals can be considered as focusing the detection of interferometric spectra only to a certain coarse depth region (see the blue axis and box in Fig. 6.1). Consequently, only a single reference arm reflector is required, further reducing the complexity of the optical arrangement. To elaborate, Fig. 6.3a describes how unwanted reflections are filtered out by a time-gate, in a scenario where multiple reflections arise from structures along the optical path. Important

parameters affecting the efficiency of the gating technique are the instrument response function (IRF) of the detector, the rise-time and uncertainty of the gate position, and the distance between the reflectors to be distinguished (Fig. 6.3b).

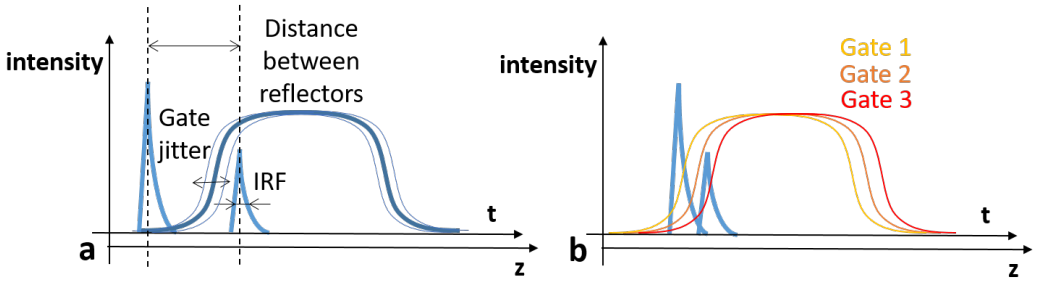

Figure 6.3: Time-gating can be applied to prevent the detection of unwanted reflections. The most important parameters determining the efficiency of the gating method are: the distance between the reflections of interest and the unwanted signal, the instrument response function, and the steepness and jitter of the gate window (a). When the reflections are too close to each other, a gate window with finite steepness either lets a significant amount of light detected from the unwanted reflection ('Gate 1'), or cuts out part of the signal of interest as well ('Gate 3') (b).

Due to the detection uncertainty expressed by the IRF, the detection of photons from a single reflector is dispersed in time. The distribution of photons in time, in a continuous case, is given by their probability density function. Since no detector can measure the arrival time with infinite accuracy, a discrete approach is used instead. The amount of photons detected at a certain narrow time region (time point with finite width) can be modelled using binomial distribution. In the case when the probability for a single photon to be detected at this time region is low and many photons arrive from the reflector, the number of photons $N_{IRF}(t_d)$ detected at a narrow region around time point t_d can be approximated using Poisson distribution, as a limit of the binomial distribution (equation 6.1). It is exactly the parameter of this Poisson distribution (μ_{IRF}) that is estimated for each time point (with finite width) when the IRF is recorded. When two reflectors are close to each other, some photons from the first reflector might be detected during the time when photons from the second reflector are expected, due to the detection uncertainty. The overlapping signals from the two closely spaced reflectors are embedded in the interferometric fringes and the A-scans, therefore the final image shows some signal from both of the reflectors. Reflectors closer to each other than the IRF, therefore, cannot be perfectly distinguished. The expected amount of overlap can be estimated using the two IRFs.

Another important case to discuss is when a reflector is close to the position of the time-gate (Fig. 6.3b). Temporal discrimination (either performed by masking the counters or gating the SPADs) has essentially a binary state. When gating (or masking)

the detected signal, each photon hitting the sensor is either counted or not, depending on its time of arrival. The gate profile in function of time is a statistical representation of the amount of gated photons $N_{gated}(t_d)$ at each time point t_d. With this model in mind and continuing to use discrete time points, the gate profile describes probability parameters $p(t_d)$ of a Bernoulli trial of $N(t_d)$ photons arriving at t_d (see equation 6.2). Gating the tail of an IRF hence becomes a Bernoulli sampling process of the photon population at t_d. With no gating, as discussed above, the number of photons at time point t_d is Poisson distributed around the expected value $\mu_{IRF}(t_d)$, described by the IRF (recall equation 6.1). It can be shown, that binomial selection of elements of a population with Poisson distributed size results in a Poisson distributed number of selected elements. The expected value of the number of selected elements is the product of the probability parameter of the binomial selection and the expected value of the original population size (equation 6.3).

$$N_{IRF}(t_d) \sim Poi\{\mu_{IRF}(t_d)\} \tag{6.1}$$

$$N_{gated}(t_d) \sim B\{p(t_d), N(t_d)\} \tag{6.2}$$

$$N_{gated}(t_d)|N(t_d) = N_{gated}(t_d)|N_{IRF}(t_d) \sim Poi\{p(t_d) \cdot \mu_{IRF}(t_d)\} \tag{6.3}$$

According to the above discussed model, when an optical signal partially overlaps with the gate, photons at the edge of the gate are accepted on a statistical basis, but with a large photon population the edge of the gate acts like an attenuator. The amount of attenuation depends on the profile of the gate edge. Due to this attenuation, closely spaced reflections may not be separated perfectly by the gate (as opposed to the situation when the gate has a perfectly sharp edge). This, in itself, is not a major limitation of the gating technique. To minimise the contribution of an unwanted reflection to the gated signal, one could, for example, perform subsequent measurements with slightly shifted gate positions, and reconstruct an estimation of the reflection of interest using a weighted sum of the gated measurements. Provided a high resolution of the gate position, analysis of the gate edge uncertainty could also be performed by calculating the difference of the gated signals acquired with neighbouring positions of the gate.

The challenge arises when gating is considered at each pixel of the spectrum. When the gate profile is the same at each pixel, a constant attenuation of the entire spectrum is provided (Fig. 6.4a). In this case, the intensity of the interference spectrum, belonging to the feature partially overlapping with the gate, gets weaker, but the shape of the does not change. Consequently, partially gated features should appear fainter in the OCT images. In contrast, mismatch of the gate position across the pixels and different IRFs (see Fig. 6.5) infer a pixel dependent attenuation (Fig. 6.4b), and causes

distortion of the gated fringes. A-scans of the distorted fringes have an increased noise level. In summary, the noise level of A-scans increases if the gates and/or IRFs are different across the pixels.

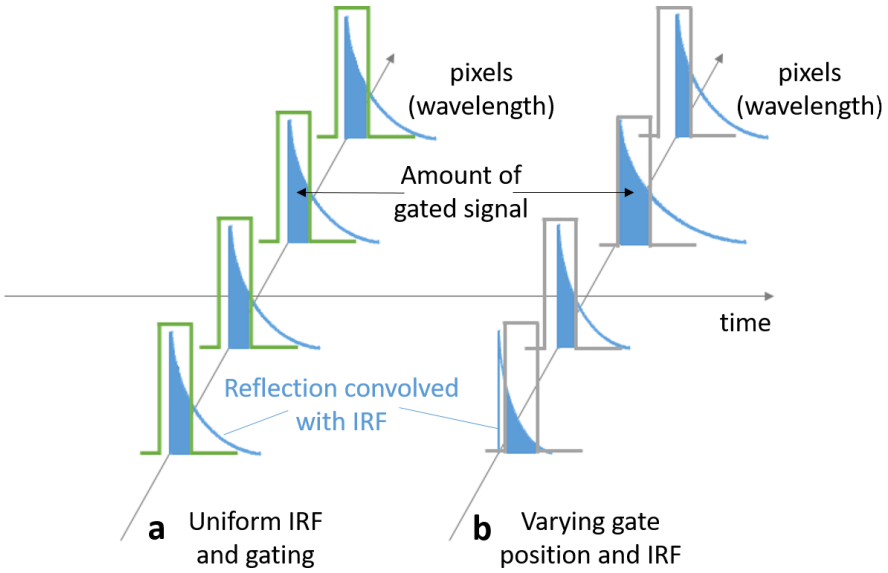

Figure 6.4: The amount of gated signal is constant across the spectrum if all the pixels have the same instrument response function (IRF) and there is no mismatch in the gate positions across the pixels (a). In reality, the IRF varies from pixel to pixel, and the gate positions are also slightly different at each pixel (b). Consequently, the amount of gated signal has variations across the spectrum as well, adding spectral noise to the gated spectrum.

A slightly different approach of temporal sectioning is to record the entire time-profile of the signals, and focus on a specific time region of interest only. In practice, this is achieved by recording the time-resolved spectral histograms and selecting certain time-bins only. This method is denoted as windowing. In this case, we face a similar problem when the window is overlapping with an optical signal. The problem again is a pixel dependent IRF and timing. However, in this case it is not the timing of the gate position that matters but the different timing of photons at each pixel. As an example, Fig. 6.5 shows the IRFs measured for red SPADs of the Ra-II sensor, using the filtered light of the supercontinuum light source. The plot illustrates well, that timing is not uniform across the pixels. Ultimately, timing mismatches across the pixels are caused by different uncertainties of photon detection (SPADs jitter), non-uniform behaviour of the time-to-digital converters (TDCs) (and delay lines, in the case of Ra-I) and potential mismatches in clock trees routing their signals to the pixels.

In the next sections removal of unwanted reflections through time-resolved detection in an OCT setup is demonstrated. First, simulations are presented to illustrate the effect of timing mismatch on the noise of the A-scans from gated measurements.

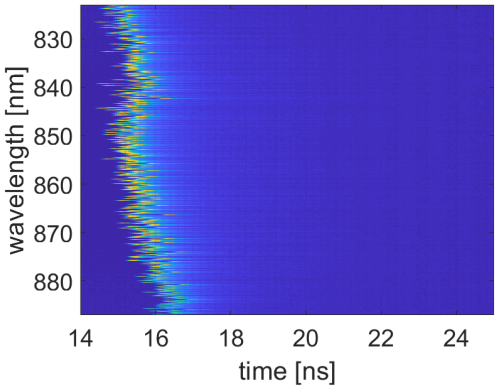

Figure 6.5: Histogram of the instrument response function (IRF) of the Ra-II red SPADs showing variations across the pixels in the timing and the IRF shape. The filtered and attenuated light of the supercontinuum light source was directly pointed at the sensor for the measurement. An approximated bin resolution of 50 ps per time-bin was used for plotting.

Following this, time-masking measurements of interference spectra are carried out both with Ra-I and Ra-II, at various distances between the source of the unwanted reflection and the sample of interest. The histogram mode of Ra-II was applied for resolving the entire time-profile of the interference signals. The time bin resolution and summing of bins were analysed for better understanding of the technique, in preparation for future studies in which extended depth OCT measurements can be performed. The chapter closes discussing these results, potential application of the time-resolved interferometric technique, and associated challenges.

6.2 Methods

6.2.1 Timing uniformity

Simulations were carried out to demonstrate how timing variations across the pixels affect the A-scans calculated from time-gated interference spectra. For this, an interference spectrum was simulated first (Fig. 6.6a), over 512 spectral points, assuming a slight path difference between two arms of an interferometer and a perfect Gaussian-shaped excitation beam. In the next step, it was simulated how the time-resolved line sensor would detect this interference signal. The interference signal is expected to be a sharp pulse in time, however the sensor sees it widened, owing to the IRF. Two spectral IRF histograms were generated. The first one was taken from a real IRF measurement (Fig. 6.6b), with each pixel's time-trace scaled to have a unity area. The second spectral IRF was meant to describe a case when timing is uniform across the pixels. This was simulated by repeating a single pixel's IRF from the first (measured)

histogram for all the pixels of the second (ideal) histogram, to construct a uniform spectral IRF. Following this, each line of time bins of the two histograms was multiplied by the original interference spectrum, to express how much of the interference signal would be detected in the respective time bins. To account for noise, Poisson random numbers were generated around the values at each pixel and each bin of the two spectral histograms. This is shown in Fig. 6.6c,d for the real (non-uniform) and the ideal (uniform) case. A window at 14.5 to 16.5 ns was defined around the peak, and the lines of time bins within the window were summed, for both histograms. In the last step, the A-scans from the two summed lines were calculated, using the previously described processing methods. Details of the simulation of the incident interferometric spectrum can be found in section C.2 of Appendix C.

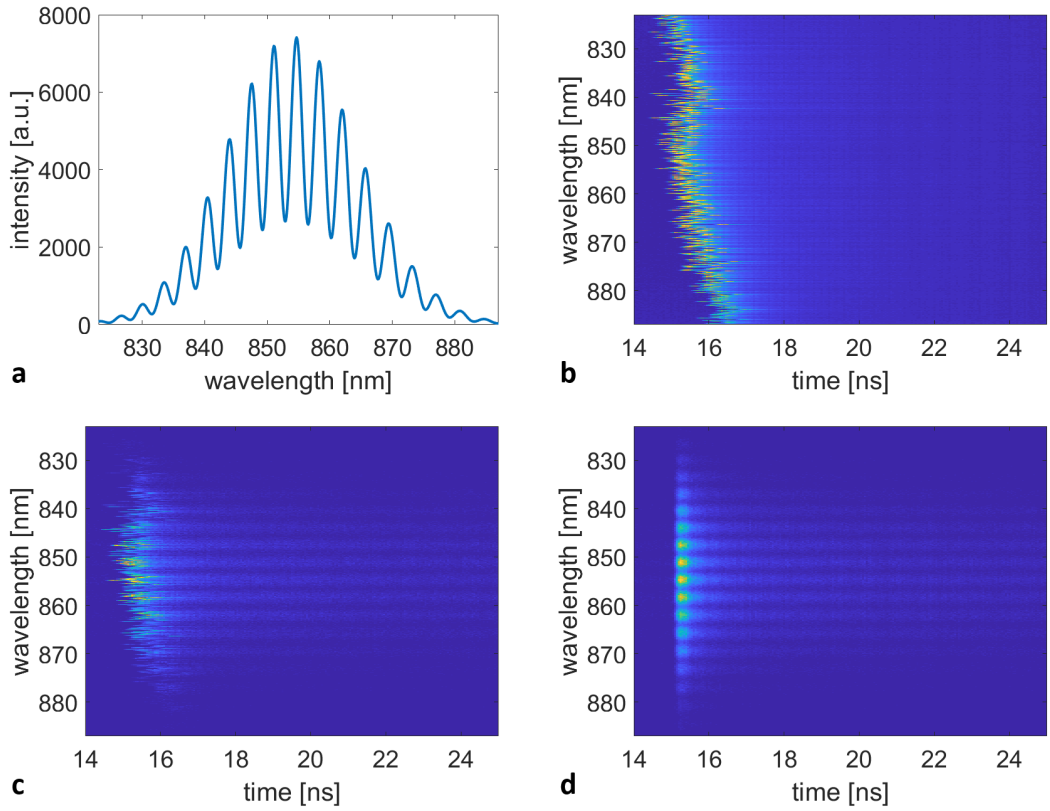

Figure 6.6: To analyse the effect of non-uniform timing across the pixels, interference fringes are simulated first (a). Spectral instrument response function (IRF) histograms are used to describe the temporal profile of the measured interference signal. The real timing behaviour of the Ra-II sensor is expressed by a real spectral IRF measurement (b). The ideal case, with uniform timing, is simulated by using the measured IRF of a single pixel, for all the pixels of the simulated, ideal case (not shown). Multiplying the two IRF histograms (at each line of bins) with the interference spectrum gives the expected number of photons detected at each time bin. Poisson distributed photon counts were generated at each pixel and bin of the temporally and spectrally resolved interference signals, both in the real (a) and the ideal (b) cases.

6.2.2 Time-gated spectral-domain optical coherence tomography

Measurements were carried with both Ra-I and Ra-II to test the removal of reflections along the beam path using temporal discrimination. In the proof of concept study, the gating performance was investigated in function of the distance between the signal of interest and the reflection to be removed. For this, a thin (~220 µm thick) glass slide was placed in the sample arm, in order to mimic strong, unwanted reflections that we would want to remove using a temporal gate/mask (Fig. 6.7). Interfering reflections from the two sides of the glass appears as two horizontal lines in the OCT images (see Fig. 6.13 to 6.15). The side of the glass slide which is closer to the fibre tip appears at zero depth in the images. The other side of the glass appears at a depth that is governed by the thickness of the glass slide. This way the actual position of the glass slide along the sample arm does not affect the image of the reflections from the slide. Several measurements were carried out at different positions of the glass slide along the sample arm. With the Ra-I sensor, the objective lens had a focal length of 30 mm (AC254-030-A-ML, Thorlabs Ltd, UK). The glass slide was placed between the fibre coupler and the objective lens. In the case of Ra-II, the focal length of the lens was 50 mm. The glass slide was placed at several positions at both sides of the lens.

As discussed above, one of the key parameters affecting the efficiency of this gating technique is the uncertainty of the detection timing. This is mostly determined by the laser pulse width and the sensor IRF. For this reason, the blue SPADs were used, with shorter IRF tail. Temporal discrimination of unwanted reflections with Ra-I was performed by enabling/disabling the photon counters of the sensor. The delay lines on Ra-I, defining the width and the position of the temporal mask, have a resolution of 1.4 ns. For finer alignment of the mask position, the laser sync signal (triggering the mask) was delayed through a delay box with 500 ps resolution (DB64, Stanford Research Systems, USA). The excess bias voltage was 700 mV. For each A-scan an exposure time of 200 µs was used.

With Ra-II, both masking the photon counters and gating the SPADs through bias modulation are possible. Experiments for temporal discrimination in SD OCT were carried out using both techniques. The excess bias voltage was 400 mV, and a single line of 1 ms exposure was captured for each A-scan. The delay box was used for crude timing of the laser sync delay. Fine alignment was accomplished using the Ra-II delay lines of 50 ps.

Two subsequent scans were performed in each measurement, one with no masking applied, and one with the temporal mask/gate turned on. The previously described data processing techniques were used to generate the OCT B-scans from the measured spectra.

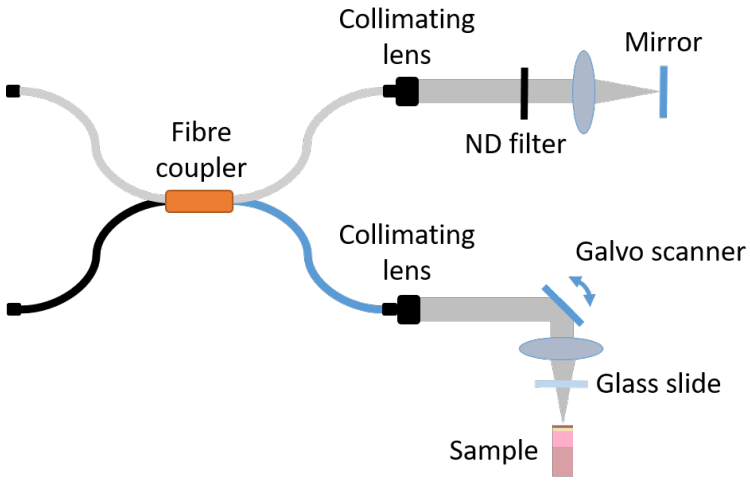

Figure 6.7: Optical arrangement in the two arms of the interferometer. A thin (~220 µm) glass slide was placed along the sample arm path to mimic strong reflections in the A-scans, independently from the actual distance of the target sample and the glass slide. The glass slide was either placed between the objective lens and the sample, or on the other side of the lens for increased distance from the sample.

6.2.3 Time-resolved spectral-domain optical coherence tomography

Time-correlated single-photon counting (TCSPC) can be applied to record the entire time-trace of interferometric spectra. Due to the slowness of TCSPC, a stable sample is needed to reduce motion artefacts. The reasonably wide IRF and low TDC resolution of Ra-I impose further difficulties in studying the time profile of the interference spectra. Therefore, histogram mode of Ra-II was used to acquire the full time-resolved spectral data, with much higher photon count rates (up to the repetition rate of the laser). In this mode, the TDCs are reset after each laser sync pulse. Photon counters, forming the time-bins with user defined resolution, are increased with the detected photons automatically.

Counts versus time and wavelength are recorded for each scanning point (top line of Fig. 6.8). At this stage, the signal can be manipulated based on its temporal content. E.g. certain time bins can be selected, related to the depth region of interest, or each line of bins (belonging to each coarse depth regions) can be processed separately. Spectral lines, similar to those given by SPC, are retrieved from the lines of bins selected (middle line of Fig. 6.8). Finally, the spectral lines of the scanning positions are transformed to A-scans (and the final OCT image), as described previously (bottom image of Fig. 6.8). For the background subtraction step of the processing chain, spectrally and temporally resolved data cubes were recorded in a similar fashion, but only with the reference arm enabled. The final OCT image in Fig. 6.8 is from an actual measurement of a fingertip, using the above described steps. The quality of the image

is in good agreement with those captured in SPC mode. The additional benefit is that time-resolved data acquisition offers means for altering the data before generating the spectral lines, based on the time of arrival.

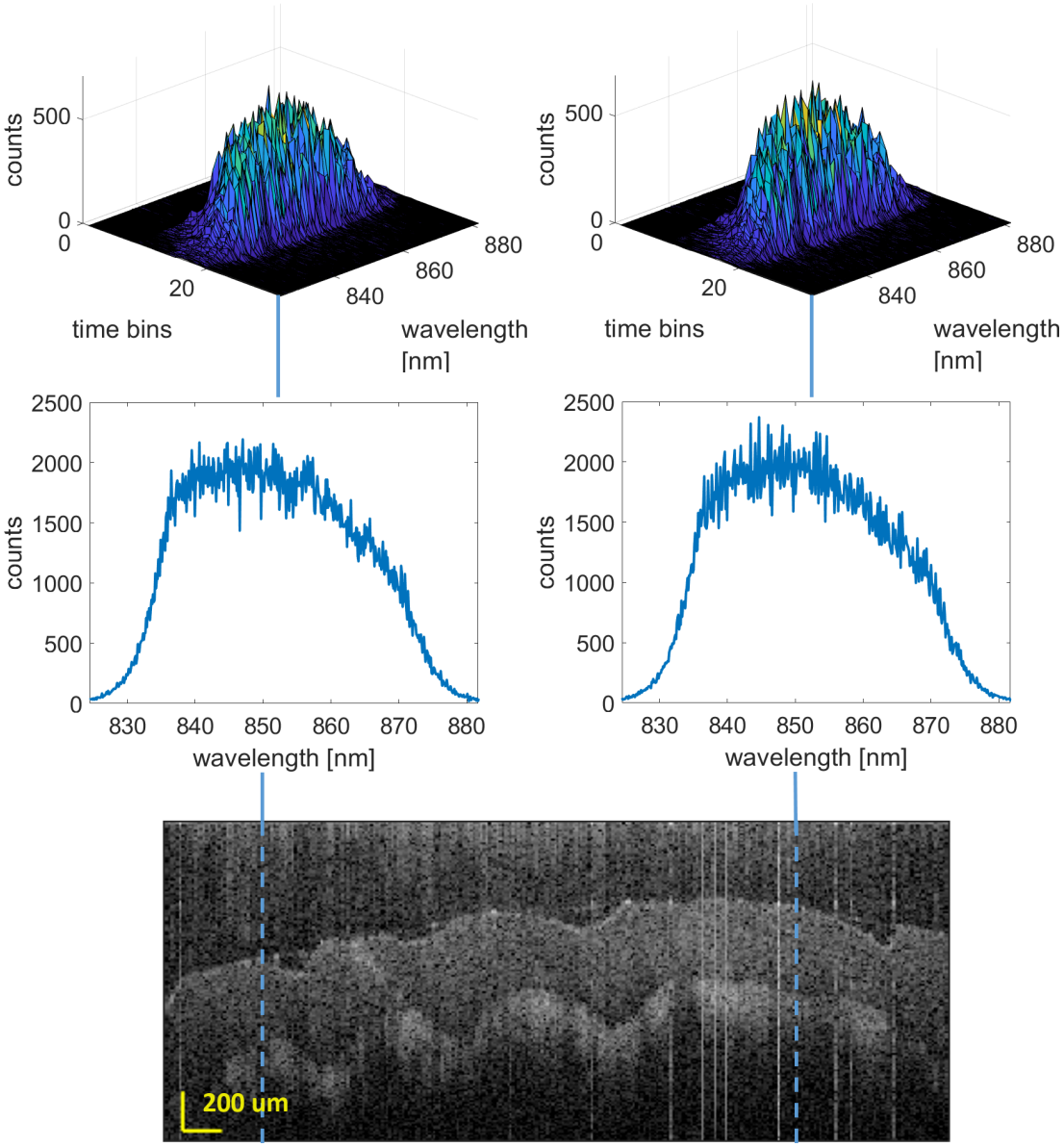

Figure 6.8: Time-resolved spectral-domain (SD) optical coherence tomography (OCT) can be applied to show a specific depth region of interest in an OCT image (or B-scan). For this, time-resolved spectral data is recorded at each scanning points (top line). The selected bin (or several bins summed) belonging to the depth region of interest provides the interference spectra at each scanning point (middle line). These spectra are then processed using standard SD OCT processing steps to generate the B-scan. In this demonstration each time-resolved spectrum was summed along the entire time axis, therefore the final image is expected to be the same as in SPC mode.

Histogram mode of the Ra-II sensor allows a zooming capability by selecting the

resolution of each time-bin (Erdogan *et al.*, 2019). At high bin resolution the reflections overlap several bins, due to the IRF. Consequently, each bin holds a lower number of counts. At low bin resolution the time-sectioning is compromised, but the number of detected counts in each time-bin is increased. An increased number of counts favours the signal to shot noise ratio. Therefore there is a trade-off between temporal sectioning capability and signal-to-noise ratio (SNR), determined by the bin width, which in turn is set by the selected histogram mode. The thin glass slide at $<$10 cm from a fixed sample was used to analyse different histogram mode settings, and compare their effect. The sample was a microscope slide, with three layers of paraffin on it. As discussed above, the interference of the reflections from the two sides of the thin glass slide appears as lines at fixed depths in the OCT image (Fig. 6.9).

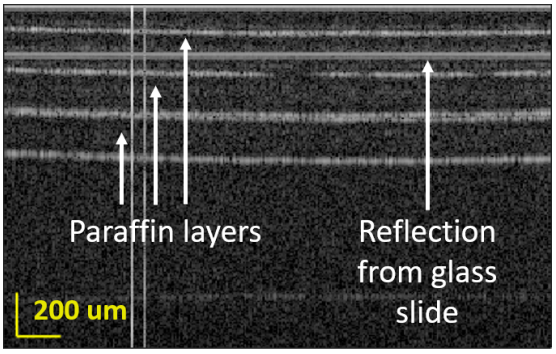

Figure 6.9: OCT image of three layers of paraffin on a microscope slide and a thin (~220 µm) glass slide at $<$10 cm distance. Different histogram modes are to be tested to see if the glass slide reflection can be removed from the image through windowing the histogram, while maintaining sufficient signal-to-noise ratio.

Various histogram modes were tested with the above described sample configuration. Spectral lines were formed by summing several bins containing the signal of interest. In each A-scan, calculated from background subtracted spectral lines, the intensity belonging to the sample's top surface was measured against the highest intensity not belonging to any feature (noise). The calculated SNR values, therefore, express the dynamic range (DR) of the images. SNR values were averaged over the scanning points, and the calculation process was repeated with different exposure times from 0.5 to 2.5 ms. The SNR (DR) values achieved in different histogram modes were compared, along with measurements in SPC mode.

As discussed, at high bin resolutions the detected optical signals from a certain reflector may overlap several time-bins (see the plots in the top line of Fig. 6.8). In these cases, summing several selected bins together increases the number of counts in the resulting spectral line, and the signal to shot noise ratio. On the other hand, when several bins with minimal or no signal are included in the summation, the noise from dark counts is unnecessarily increased in the A-scans and the final image (Fig. 6.10). The

effect of the number of selected time-bins on the achievable sensitivity was measured using two mirrors in the interferometer with a slight path difference. The sample arm signal was attenuated to a level where only a minor interference signal was detected by the sensor. The interference signal was recorded using the histogram mode with the highest, 50.4 ps, bin resolution, which corresponds to a step size of 7.5 mm along the depth. Spectral lines were generated from the time-resolved data by selecting a progressively higher number of time-bins around the bin with peak intensity. A-scans were generated from the background subtracted spectral lines. The peak in each A-scan belonging to the signal from the mirror was measured against the highest peak belonging to noise. The measured attenuation of the neutral density (ND) filters was added to the SNR for the final sensitivity values. The measurement was repeated several times in order to minimise statistical uncertainties. The exposure time was 1 ms in all cases.

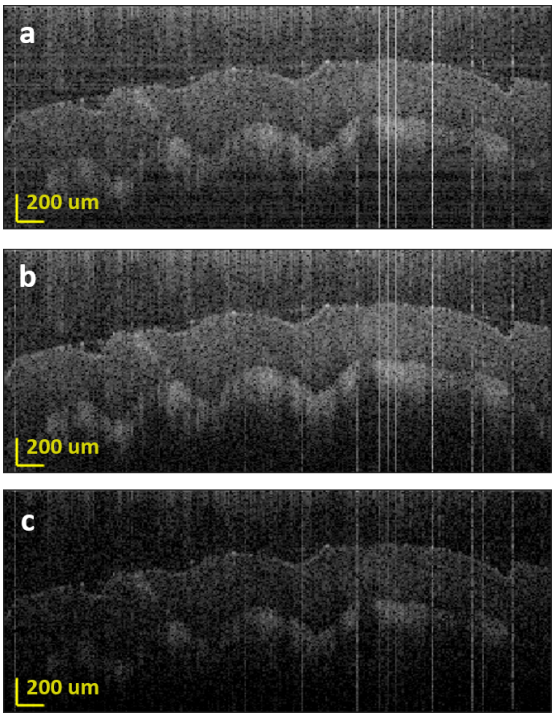

Figure 6.10: Three OCT images of a fingertip acquired in time-resolved (histogram) mode of Ra-II, demonstrating the effect of the number of time bins selected for spectral lines. When an optical signal overlaps several time bins, summing multiple bins increases the number of counts in the extracted spectral lines, which in turn increases the signal-to-noise ratio of the A-scans. Summing too many bins (with low signal) however amplifies the noise. The noise is high in the image where 24 time-bins were summed for the spectral lines (a). The noise is reduced but the signal intensity is maintained when only 8 bins are selected (b). If only a single time-bin is used, the low number of counts in the spectral lines compromise the signal-to-noise ratio and the image quality (c).

Temporal sectioning experiments using the histogram mode of Ra-II were also performed in a similar scenario as with gating/masking. For convenience, a fingertip was used as a sample and a thin glass slide at roughly 4 cm distance from the fingertip. The actual sample (fingertip) in this case is less relevant, since the focus is on the reflection from the glass slide, which is to be removed through gating.

The selected time-bin resolution was 50.4 ps. 7 bins of the time-resolved data were summed for spectral lines. These bins included the signal from the fingertip only. The exposure time was set to 1 ms. The objective lens used had a focal length of 30 mm and the scanning range was 2.5 mm. Given the bin resolution (corresponding to 7.6 mm distance), the applied optics is not expected to effect the temporal sectioning. For background lines, the same 7 time-bins were summed on the time-resolved data from the reference arm.

6.3 Results

6.3.1 Timing uniformity

The simulation results show a difference of 9.8 dB in the baseline on the A-scans (calculated as an average of FFT bins 28 to 256) (Fig. 6.11). The reason for an increased base line (noise level) when the true IRF is used is mismatched windowing, and therefore varying attenuation of the windowed spectrum at each pixel. In contrast, Fig. 6.12 shows the A-scans calculated from a uniformly windowed interference spectrum, and the spectrum when no windowing is used, but the entire time-trace is summed. In this case, we see an attenuation of the peak, since the tail of time-resolved signal is clipped by the window. However, because the window cuts the signal uniformly, no noise increase is observed.

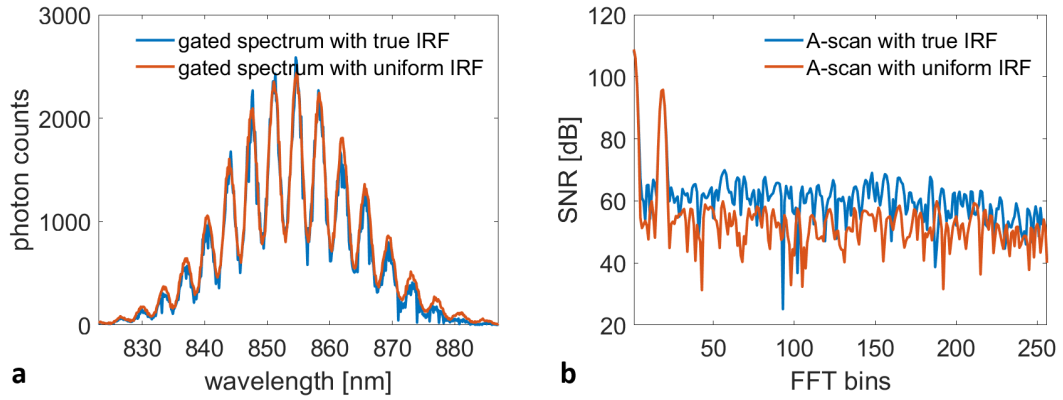

Figure 6.11: Windowed interference spectra from the simulated time-resolved spectral histograms (a), and the respective A-scans (b).

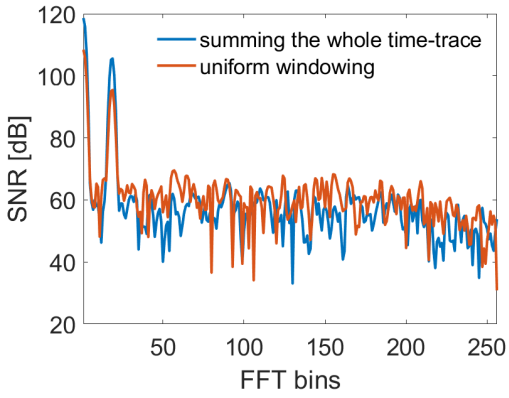

Figure 6.12: The A-scan peak decreases if not the entire signal is collected within the time-window, but provided that the cutting is uniform across the pixels, the interference spectrum will not be distorted, adding no noise to the A-scan.

6.3.2 Time-gated spectral-domain optical coherence tomography

Figure 6.13a shows the image of the fingertip with the additional glass slide, taken with Ra-I in SPC mode without any temporal discrimination. Straight after taking the non-masked image, another image was acquired with the temporal mask turned on (Fig. 6.13b). A randomly selected A-scan is shown in Fig. 6.13c without and with masking applied. The intensity was suppressed by 23.3 dB in the case of the reflection from one side of the artefactual glass slide (zero depth), and 18.7 dB for the reflection from the other side (~220 µm depth). The temporal mask position and width have a low resolution with Ra-I. The mask position resolution is 500 ps, set by the external delay box, and the mask length resolution is 1.4 ns, set by the delay line resolution of Ra-I. The timing characteristics and long IRF of the sensor required that the reflection arises from no shorter distance than 10 cm from the sample.

Ra-II has superior timing characteristics. This allows removal of the glass reflection using the described masking technique, even when the glass slide is only 3 cm away from the sample (Fig. 6.14a,b, without and with masking, respectively). Owing to the short IRF, the signal from the glass slide does not overlap with the signal from the sample, hence it can be removed entirely. Compared to the case with Ra-I, the OCT image is noisier when the mask is turned on (Fig. 6.14b). The most probable reason for this is an overlap between the glass reflection's IRF tail and the gate, which varies somewhat from pixel to pixel, and/or a relative mismatch in the gate lengths per pixel, see Fig. 6.4 and section 6.3.1.

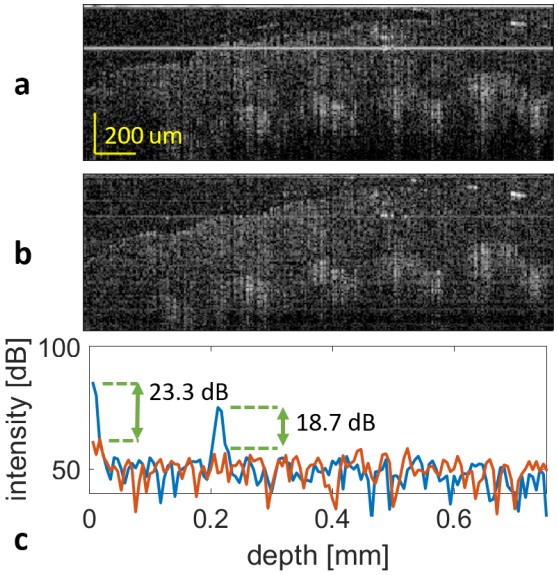

Figure 6.13: Image of a fingertip and glass slide without (a) and with temporal discrimination (b). Reduced reflection from the glass slide is shown in a single A-scan (c). The exposure time was 200 µs. The distance of the glass slide from the fingertip was 10 cm, but the interfering reflections from the sides of the glass slide (common path interference) are naturally superimposed on the fingertip depth profile.

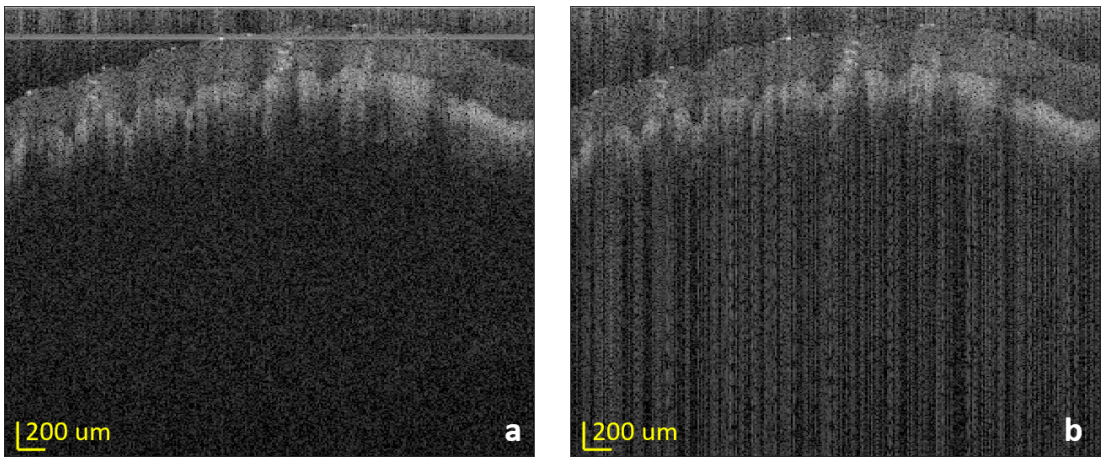

Figure 6.14: Ra-II allows complete removal of the glass slide reflection at a distance of 3 cm from the fingertip using the masking technique. Both the fingertip and the glass slide are visible in the image with no masking applied (a). With masking, the noise in the image increases due to per pixel mask variation and clipping of the tail of the glass reflection (see section 6.3.1 and Fig. 6.4) (b). The exposure time was 1 ms.

When the distance is even shorter between the reflectors, the signals from the sample and glass slide are more and more overlapping, due to the IRFs, obstructing complete removal of the glass reflections. Figure 6.15 demonstrates this case, when the distance between the fingertip and the glass slide was 8 mm. The rising edge of the mask was set between the glass and fingertip signals, and scanned with the smallest possible step size (50.4 ps), as shown in Fig. 6.15d. Three particular cases are depicted with 100.8 ps difference in the mask edge position between them (Fig. 6.15a to c). As seen in the masked images (Fig. 6.15a2,b2 and c2), the glass reflection is getting fainter as the mask is moving away from the signal of the glass slide. However, as the mask moves away from the glass reflection, the signal of interest also decreases as a decreasing fraction of it is covered by the gate.

The Ra-II sensor allows gating of the blue SPADs through modulating their bias voltage. As discussed in chapter 3, in the case of the gating method, the timing of the gate edges have more strict requirements compared to the counter masking method. Namely, a minimum time delay has to be maintained between the rising edge of the SPAD enable signal and the falling edge of the QUENCH signal (see Fig. 3.7). The advantage of the gating method compared to masking is that the SPADs are completely turned off in the 'off' period, making them insensitive to strong, saturating optical signals. Therefore, a faint signal of interest can be detected, even in the presence of saturating reflections. For this, the strong optical signal has to arrive in the time region when the SPAD is disabled, and the signal of interest has to arrive when the SPAD is enabled and the QUENCH signal is low (Fig. 6.16a). With the current design of Ra-II, these time-regions have to be several ns apart. Thus, this gating method is impractical if the source of the saturating signal is within a few cm vicinity of the structure of interest.

Measurements were performed for gating off glass slide reflections from a few cm distance of the sample of interest. In these cases the unwanted signal arrived in the 'on' period of the SPADs, but when the QUENCH signal was still high (Fig. 6.16b). The signal of interest arrived when the SPADs were still enabled, but the QUENCH signal was already low. No successful gating could be performed in these cases, indicating that a saturating signal hitting an enabled SPAD (even if arriving when the SPAD is quenched) prevents the detection of the signal of interest.

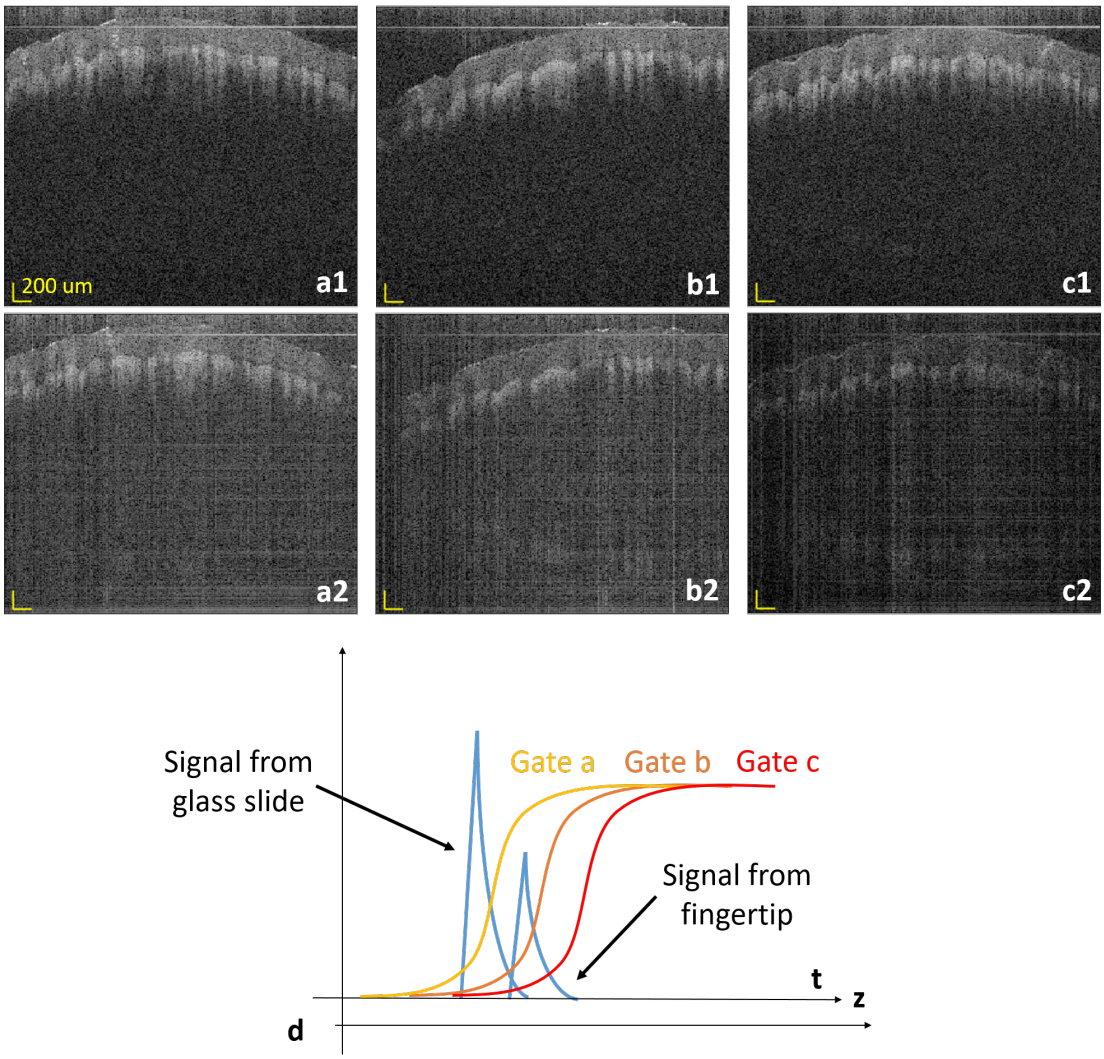

Figure 6.15: Non-masked/masked image pairs with 100 ps (corresponding to 15 mm) between the mask positions (a1,a2; b1,b2 and c1,c2). The glass slide from which the reflection is to be removed is at 8 mm distance from the target. At this distance, the signals convolved with the instrument response function are overlapping. Depending on the mask position, either the glass slide reflection is not gated entirely (a2), or the signal from the fingertip is masked off as well (c2). When the gate cuts either reflection (signal from the fingertip or the glass slide), the gate position variation across pixels adds varying amount of noise to each A-scan. The exposure time was 1 ms in each case.

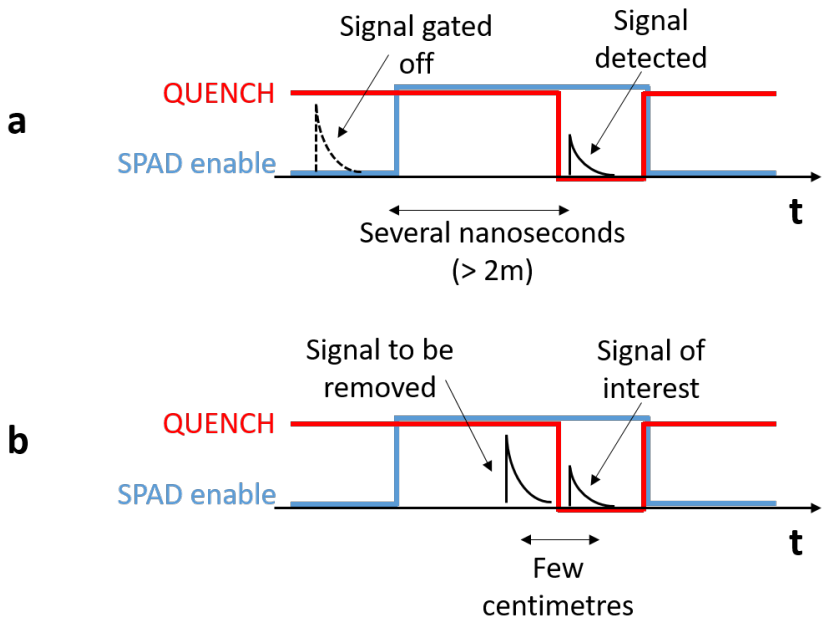

Figure 6.16: Timing of the SPAD gating signals of Ra-II. For complete removal of a reflection, it needs to be several nanoseconds ($>$2 m) away from the signal of interest (a). If the distance between the signal of interest and the unwanted reflection is only a few centimetres, the unwanted reflection arrives to the SPADs while they are enabled (b).

Time-resolved spectral-domain optical coherence tomography

Unwanted reflections (such as the one from the thin glass slide) can only be removed from the OCT image if they are well separated from the signal belonging to the sample (microscope slide). The effect of the time-bin resolutions is shown in Fig. 6.17, for a bin resolution of 201.6 ps (Fig. 6.17a,c) and 352.8 ps (Fig. 6.17b,d). Respective B-scans are shown in Fig. 6.17c and Fig. 6.17d.

SNR of the top layer of the paraffin sample is shown in Fig. 6.18, when different bin resolutions are used. Even though the measured SNR values have statistical variations, generally they follow the same exposure time dependency as the SNR in SPC. Two major factors are considered governing the SNR, the amount of signal detected (with its the respective amount of shot noise) and dark counts. Several bins were summed in each case to generate the interference spectra. The range of bins to sum was determined to cover all the bins over which the photon counts were dispersed due to the IRF. As a result, the number of counts in the interference spectra was the same in each case. Not considering outliers, the SNR values of different histogram modes are reasonably close to each other. This indicates that dark noise is also similar in each case, even though there is a doubling of the bin size in subsequent histogram modes tested. This is plausible given the dark count rate (DCR) of SPADs

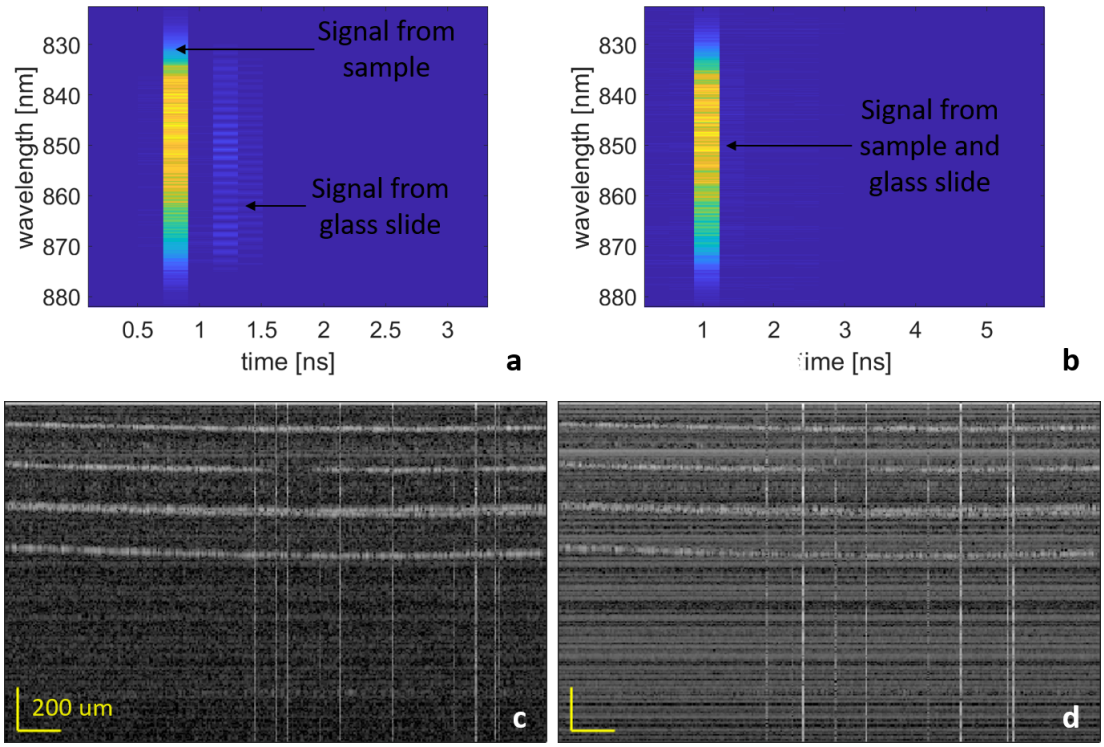

Figure 6.17: With 201.6 ps per time-bin resolution, the reflections from the microscope and glass slides are well separated (a), and therefore the unwanted reflection from the thin glass slide can be easily removed from the B-scan using the described windowing technique (c). At a lower bin resolution of 352.8 ps, most of the reflection from the thin glass slide falls to the same time-bin as the signal from the sample (b), and hence cannot be entirely removed from the B-scan (d).

and the short length of exposure time. Cases marked with circles in Fig. 6.18 could not distinguish between the signal of interest and the unwanted reflection.

Figure 6.19 shows the measured OCT sensitivity in function of the number of bins summed for generating spectral lines. Selecting only a few time-bins gives lower counts in the generated spectral lines, and a lower sensitivity consequently. This is because the signal of interest extends to several time-bins, since the IRF is wider than a single time-bin of 50.4 ps. Above 8 bins, the sensitivity does not increase further, since the additional bins do not contain any signal. We see no significant decrease in sensitivity either, even though the additional time-bins only include noise. The reasons for this is twofold. On one hand, the dark count contribution of a single bin (compared to dark counts in SPC mode) is expected to be minor, since a single time-bin of ~50 ps covers only $\frac{1}{1000}$ of the 50 ns laser period. On the other hand, DCR noise has minor contribution on the SNR at this exposure time in the first place, as seen before. Indeed, the highest sensitivity (86 dB) is similar to that achieved in SPC mode, again indicating that the contribution of DCR to the overall noise is secondary.

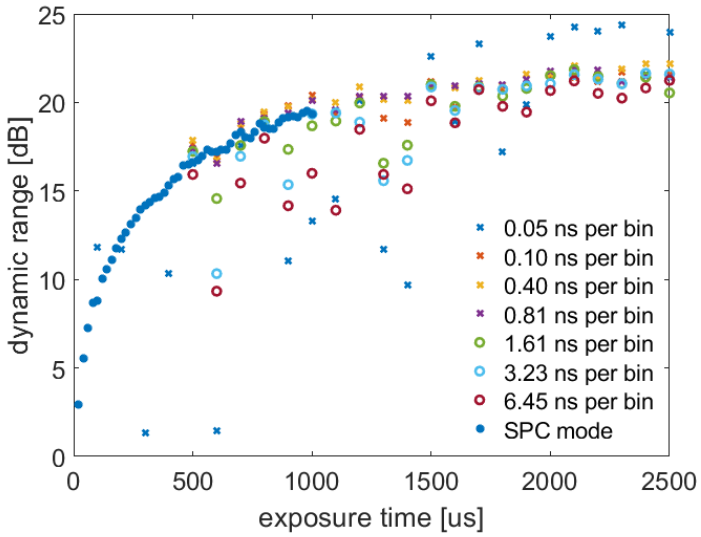

Figure 6.18: Dynamic range (DR) of optical coherence tomography images in function of exposure time, acquired in time-resolved mode with varying bin resolution, and single photon counting. The sample was three layers of paraffin on a microscope slide. An additional glass slide was placed $<$10 cm away from the target to test the effect of bin resolution on temporal-sectioning. DR was calculated as the signal intensity of the top surface of the sample compared to the highest noise intensity, averaged over the A-scans. DR of time-resolved detection in function of exposure time generally follows the trend of DR in single-photon counting mode. Measurements marked with circles cannot differentiate between the sample and the glass slide because both signals fall to the same time bin.

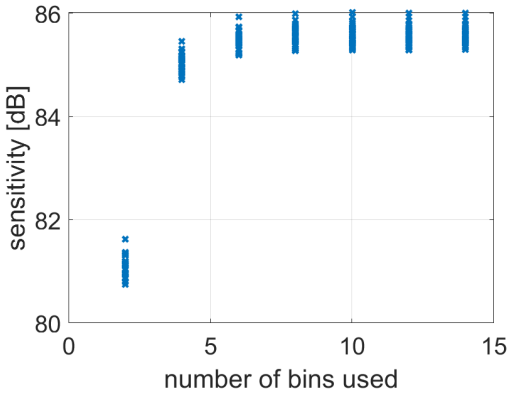

Figure 6.19: Sensitivity of the optical coherence tomography system using the histogram mode of Ra-II with 50.4 ps bin resolution, in function of the number of bins selected for generating spectral lines. Several bins needs to be summed, since the instrument response function is wider than a single time-bin. The highest sensitivity lines up with single-photon counting measurements, and stays constant above 8 time-bins, indicating that dark count rate has minor effect on sensitivity. The exposure time was 1 ms. 37 measurements are plotted to show statistical variations.

Finally, Fig. 6.20 shows the effect of temporal sectioning using windowing, with a fingertip in focus of the sample arm objective and a glass slide ~4 cm away (directly placed on the objective lens). Figure 6.20a shows the image acquired in SPC mode, using 1024 pixels. In Fig. 6.20b the image taken in histogram mode is shown, with the detection focused on the signal from the fingertip only. Here, only 512 pixels could be used, explaining the shorter depth of field.

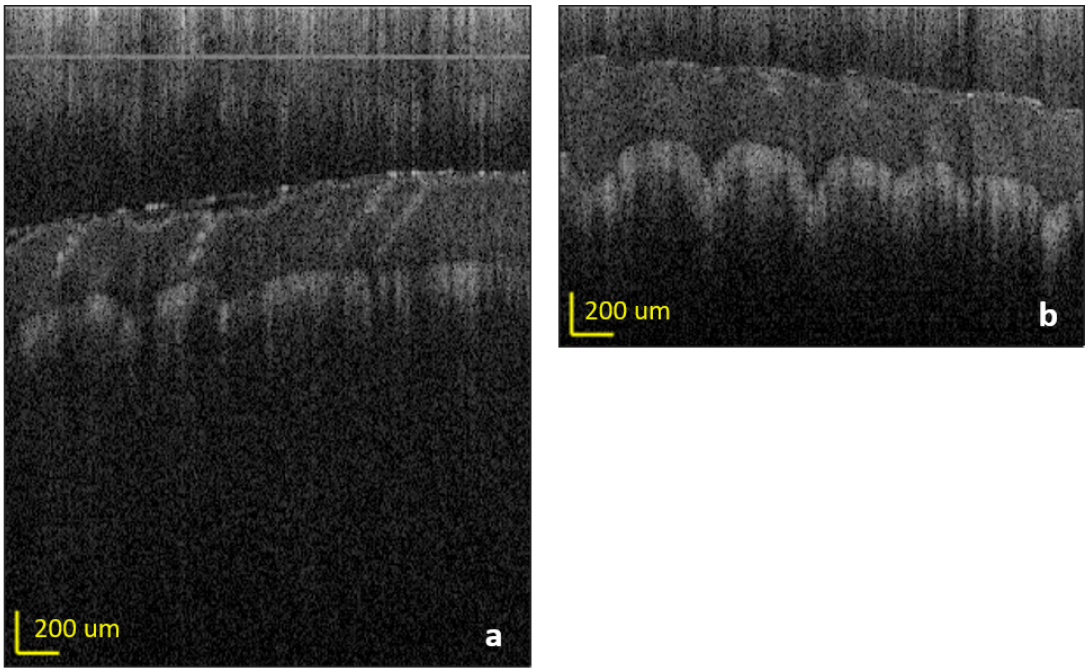

Figure 6.20: Optical coherence tomography image of a fingertip and a glass slide at a distance of roughly 4 cm (a). The glass reflection is removed by windowing the time-resolved interference spectra (b). For this, the histogram mode of Ra-II was used with a bin resolution of 50.4 ps. The exposure time was 1 ms.

6.4 Discussion

In this chapter, a novel detection scheme is proposed combining time-resolved detection and low-coherence interferometry. The overall approach is illustrated in Fig. 6.1. Assuming a multiply reflective reference arm along depths d1, d2, ..., etc., spectral interferometry can be obtained for each separate depth by deploying time-gating or TCSPC.

6.4.1 Time-gating and masking

Time-gating or windowing can be used to remove unwanted reflections and overcome saturation problems. In the presence of strong, saturating reflections, the reference arm power is adjusted, so that the reflection does not saturate the detector. Consequently, the lower limit of detection, in this case, will not be that set by shot noise but the finite dynamic range of the detector. Time-gating was shown to be effective in these scenarios to prevent detection of the strong reflection, so that the reference arm power can be set to a high level where the detection is shot noise limited, increasing the contrast of weak reflectors in the OCT images (Muller and Fraser, 2009).

Several tests were performed with the Ra-I and Ra-II line sensors to remove strong reflections from the OCT images, using time-masking. The timing characteristics of the Ra-I sensor required that the distance between the reflector to remove and the sample of interest is no shorter than 10 cm. Non-linear optics, where the primary beam is gated provide more accurate timing (Muller and Fraser, 2009), but requires high power lasers and more complex optical systems. Furthermore, optical shuttering only allows to remove a certain time region, and not complete time-resolved detection of spectra (see next section). Ra-II already achieves 8 times better timing performance than Ra-I, allowing the removal of reflections from a few cm distance. The IRF of Ra-II is also much shorter, still it compromises how well closely spaced structures can be distinguished. This is also affected by the sharpness of the mask edges. In Fig. 6.15, the distance between the sample and the unwanted reflection (8 mm), was similar to the step resolution of the gate along the depth (7.5 mm). The figure shows that a slightly shifted gate does not change the amount of gated signal abruptly, confirming that the widened temporal profile of the optical signals (due to IRF) and the finite rise of the gate edge overlaps, and prevent perfect cut of the unwanted signal.

Short IRF and uniform timing is needed to avoid spectral noise in the measured interference signal when gating is applied. In Fig. 6.13 no increased noise is observed when turning the mask on. The most likely reason for this is the low timing resolution. Due to the low resolution of the mask position in Ra-I, the reflections from the sample and the glass slide had to be relatively far from each other. At such distance the reflections were not clipped by the mask considerably. Timing mismatch across the pixels can also be introduced by inter pixel delay, due to the spectrometer configuration (slightly different path lengths for different wavelengths). The Ra-II sensor chip has a width of ~12 mm and the spectrometer's focusing lens has a focal length of 150 mm. From this, the path length of the ray hitting the edge of the sensor is about 150.12 mm. The 0.12 mm path length difference of rays reaching the sensor's centre and edge corresponds to 0.4 ps timing difference, which is negligible with respect to the time resolution of Ra-II. With different optics, sensor size and temporal resolution, this effect

may need to be considered and potentially corrected for. This could be performed, e.g. by measuring the entire time-resolved spectra, and applying a specific time window for each pixel, which would also reduce the spectral noise caused by non-uniform gating.

The main limitation of the results shown is that the temporal discrimination was carried out through masking the counters, instead of bias modulation of SPADs. While this is efficient for removing non-saturating signals from the OCT images, imaging is still not protected against saturating reflections. In order to provide complete insensitivity against reflecting signals at short distances from the weak reflecting region of interest, SPAD gating (when the SPADs are completely turned on/off) needs to have better temporal resolution.

6.4.2 Time-resolved detection

The proposed approach of combining time-resolved detection and low-coherence interferometry has potential for extended depth imaging, beyond ~2–3 mm that is usually encountered in spectral-domain optical coherence tomography (SD OCT). While this technique still needs thorough experimenting and further studies, the histogram mode of Ra-II was shown to be suitable for time-resolved detection of the interference spectra, and temporal sectioning of features along the beam path. In histogram mode, each time bin has a limited depth. Up to 1024 photons can be counted at each bin when 32 of them are used. The maximum photon count rate in this mode is equal to the repetition rate of the light source. The limited counter depth requires short enough exposure times, so that less than 1024 light pulses are accounted during the exposure. If this is still problematic e.g. with a very high repetition rate light source, the counters can be extended to 20 bit depth, operating with only 16 time bins (Erdogan *et al.*, 2019). In the histogram mode, the time-span of the histograms is restricted by the fixed number of time bins, therefore high bin resolution comes with a short observable time range, and vice versa. This affects how many coarse depth regions can be distinguished (ruled by the number of bins), and how wide they are (ruled by the bin resolution).

In contrast, TCSPC does not trade temporal resolution for the detected time range. The former is determined by resolution of the TDCs (50 ps), and the latter is set by the cycle time of the pulsed light source. The drawback of the technique comes at the acquisition time, given by the fact that TCSPC operation requires low optical powers and has a 1 photon per pixel per exposure time upper limit on detection rate. For quantitative comparison of the two acquisition methods, let us assume a 20 MHz repetition rate of the light source and a maximum photon count rate of 1 % of the source repetition rate at each pixel, for TCSPC operation. Let us further assume an exposure time that is no longer than 100 pulses of the source (5 µs) in order to prevent TCSPC artefacts (pile-

up, counting loss). A minimum acquisition time for collecting 1000 photons at a certain pixel is, therefore, 5 ms in TCSPC mode. The collection efficiency of histogram mode has a maximum of one photon per laser pulse (Erdogan *et al.*, 2019). With the same repetition rate (and a higher optical power) the minimum time to collect 1000 photons is only 50 µs, which is 100 times shorter. TCSPC is not applicable for time-resolved SD OCT if minor movements of the observed sample are expected (since this would cause fringe washout), but only on fixed samples.

The execution time of data acquisition, processing and plotting in histogram mode was measured by profiling the respective software code. The acquisition of a single OCT image of 300 A-scans, each with 1 ms exposure time was 385.10 ms in average, if only the raw data was stored. The execution time was 421.33 ms with live processing of raw data and plotting, indicating that histogram mode is a good candidate for live acquisition in practical clinical applications, especially with more efficient SPADs and, therefore, reduced exposure times. The achieved SNR at this rate (85 to 86 dB) is similar to that in SPC mode.

Short time bins and efficient photon processing are not the only requirements for efficient temporal sectioning, the detection also needs to have a short timing uncertainty. The effect of the IRF is important for extended depth OCT as well. When a signal overlaps several bins, it may appear in each of the affected depth regions and the respective fine-resolved depth profiles. Potentially, this can be mitigated by post processing the data, however this requires further considerations and detailed study of the effect.

Shorter jitter values (e.g. as reported by Burri *et al.* (2016)) are expected to enable improved temporal performance in time-resolved spectral interferometry. Other applications, such as 3D ranging (e.g. LIDAR) are likely to benefit, where surface shape is of more interest. We also believe that the platform will be useful for studying highly scattering tissue samples, as improved timing with SPADs will allow millimetre and sub-millimetre gating precision. Finally, combining common-path interferometry with time-resolved detection will be useful for both fluorescence and backscattered scenarios, in multi-modal systems.

Chapter 7

Summary and Conclusion

7.1 Thesis summary

Complementary metal-oxide-semiconductor (CMOS) single-photon avalanche diode (SPAD) sensors with high resolution timing circuits are being more and more widely applied in biophotonics applications. Bruschini *et al.* (2019) provide a thorough review of the scientific literature on CMOS SPADs in biophotonics in the last 15 years. In this thesis, the unique feature of SPAD sensors of spectrally and temporally resolved light detection on the single-photon level was applied, focusing on improvements in three techniques in particular. The three techniques: time-resolved fluorescence spectroscopy (TRFS), time-gated Raman spectroscopy and spectral-domain optical coherence tomography (SD OCT) are all associated with different light levels and they set diverse requirements towards an ideal detector. Two CMOS SPAD based linear sensor arrays with per pixel timing circuits were the key components in the research, the Ra-I sensor (Krstajić *et al.*, 2015; Kufcsák *et al.*, 2017) and its improved version, Ra-II (Erdogan *et al.*, 2019).

7.1.1 Fluorescence and Raman spectroscopy

Time-resolved fluorescence spectroscopy is a well established technique for measuring nanosecond fluorescence decays, using time-correlated single-photon counting (TCSPC) (Millar, 1996). Single-photon counting detectors, such as photomultiplier tubes (PMTs) and microchannel plates (MCPs) have been traditionally used for this, with external timing instrumentation, typically in the form of Peripheral Component Interconnect (PCI) cards (Becker *et al.*, 2002). The main limitation of this approach is the low number of detection/timing channels. SPAD arrays offer an unprecedented capability of massively parallel single-photon detection and timestamping with tens of picoseconds, on several hundred channels.

In the TRFS studies, a complete platform was built around the Ra-I sensor, including a pulsed light source, optical arrangement and sensor controlling firmware and software. The platform offers outstanding photon processing efficiency for live measurement

of spectrally and temporally resolved fluoresce kinetics, underpinned by multichannel timing and highly improved line rate (27 times higher than with the previous version of the sensor), supported by firmware and software. The upper limit of the photon processing capability was investigated in experiments where fast transients of spectral fluorescence lifetimes were measured. It was found that changes in fluorescence decays and lifetimes over nanometre wide spectral bands can be measured with tens of milliseconds sampling, over a single chemical reaction transient. Alternatively, millisecond sampling of lifetime changes can be achieved at the expense of spectral resolution. The sampling interval of spectral lifetime estimates can be reduced to a few hundred microseconds without measuring the decays, when using the centre-of-mass mode of Ra-I. The platform has great flexibility when choosing the acquisition parameters, and offers a trade-off with regards the accuracy of decays, transient sampling interval and spectral resolution. The system was developed with versatile functionality, yet easy operation in mind to support other studies and potential clinical use in the future as well. Functionality of the platform was further demonstrated in Raman spectroscopy, utilizing the time-gating feature of Ra-I. Time-gating is applied to reduce the noise of the detector and potentially reject fluorescence contribution from the sample, hence improving the measured Raman spectra.

7.1.2 Optical coherence tomography

Optical coherence tomography has greatly benefited from the introduction of Fourier-domain optical coherence tomography (OCT) techniques (de Boer *et al.*, 2017). In industrial applications the most widely used variant is spectral-domain (SD) OCT, where the depth profiles are recovered from spectral interference measurements. The traditional sensors of SD OCT systems have been charge-coupled device (CCD) based and CMOS photodiode arrays. Single-photon detectors with no readout noise were applied in a traditional, time-domain system, to achieve increased signal quality (Mohan *et al.*, 2008). For the first time, this thesis investigates the applicability of single-photon detection in spectral-domain systems. An SD OCT system was built and described, with both the Ra-I and Ra-II sensors tested. The parameters of SPAD based sensors (and those of Ra-II in particular) affecting the OCT imaging performance were studied, both through theoretical considerations, simulations and measurements. The results suggest that the most important limitations of the applied line sensor, compared to published SD OCT systems, are the detection efficiency of near-infrared photons, and the challenge to tackle strong optical signals saturating the sensor.

Using a pulsed source and the time-resolving capability of Ra-II, a novel detection scheme is introduced combining low-coherence interferometry and time-resolved detection. The interference spectrum is measured across the pixels of the line sensor,

with the photons of the spectrum time-stamped. A proof of concept study is presented, where the proposed detection scheme is applied to remove reflections from along the depth of the sample in the OCT setup. Removal of reflections from an artefactual glass slide in 3 cm distance from the sample was demonstrated by time-masking the photon counters of the sensor. Partial removal of the reflection was possible at sub-centimetre distances.

7.2 Discussion and conclusion

While SPAD based sensors have exceptional performance in many respects, it needs pointing out that they are far from being perfect. The performance that competing sensing technologies offer is often superior with respect to certain detection parameters. For example, state-of-the-art streak cameras have higher time-resolution than integrated time-to-digital converter (TDC) and time-to-analogue converter (TAC) based circuits in CMOS SPAD sensors. This, of course, comes at the price of bulky and expensive instrumentation. More importantly, the extent of parallelisation and the number of timing channels in SPAD sensors is second to none. SPAD designs tailored to specific applications have been reported in the literature, but it is clear that improved characteristics require trade-offs to be made. Sticking with the previous example, the integrated, massively parallel timing capability usually means a limited fill-factor, due to the ancillary timing circuits on the detector.

Innovative ideas, and emerging new technologies present novel solutions to overcome limitations and eliminate the need for trade-offs. To overcome the limited fill-factor, and hence low detection efficiency, various solutions have been proposed, such as using linear array formats (Krstajić *et al.*, 2015; Kufcsák *et al.*, 2017; Erdogan *et al.*, 2019), relying on external timing circuits (Burri *et al.*, 2016), or 3D stacking (Abbas *et al.*, 2016). The aim of studies like this thesis are to feed into new designs, identifying the limitations and key requirements towards the sensors in real-world applications. In the case of the applications presented in this thesis, these requirements involve higher sensitivity for near infrared (NIR) light, more efficient photon counting with pulsed sources, lower jitter and more uniform timing with higher resolution gating, as discussed in the respective chapters. It is also important to realize that the improvement of applications not only builds on novel sensors, but also benefits from other advanced technologies. These include improved light sources, field-programmable gate arrays (FPGAs), and general-purpose graphical processing units (GPGPUs).

The biggest advantage of massively parallel, time-resolved CMOS SPAD sensors in TCSPC is the ability to overcome pile-up issues, easing the low count-rate requirements of TCSPC (Arlt *et al.*, 2013). In the TRFS studies of this thesis, it was not only

the number of parallel channels that improved photon processing considerably, but the 'centre of mass' mode (CMM) as well. The CMM offers more efficient photon timing at the expense of loosing the information of the temporal shape. The ideal solution is a method that provides high photon processing and timing and yet capable of recording the full temporal shape of signals, with high resolution. Histogram mode of the Ra-II sensor serves this purpose (Erdogan *et al.*, 2019). This state-of-the-art timing method can process up to one photon per laser cycle. Its resolution can be varied from 50 ps to 12.8 ns. The only constraint of this technique is a limited number of photons per time-bin, and a limited timespan of the histograms, which is inversely proportional to the bin resolution. With the technological improvements and the significantly increased performance SPAD sensors have achieved in a decade (Bruschini *et al.*, 2019), it is expected, that efficient photon processing solutions coupled with high sensitivity will further advance fast fluorescence techniques, such as spectral double kinetic measurements or spectral fluorescence lifetime imaging microscopy (FLIM). These improvements introduce a new challenge, handling and processing huge amounts of data. The amount of information generated from strong optical signals with efficient photon counting demands high detector frame rate. Data reduction methods (e.g. histogram mode instead of TCSPC) have huge relevance too. Applications in clinical environments often require live information on the measured data (e.g. in interventional radiology). This requirement puts emphasis not only on the sensor, but also on the surrounding processing electronics. Advancements in the field of embedded systems, FPGA and digital signal processor (DSP) technology is likely to have a significant impact as well.

In other cases when the optical signals are weak (e.g. in Raman spectroscopy and single molecule localisation microscopy (SMLM)), the emphasis is shifted to low noise and high sensitivity (high photon detection efficiency (PDE)). An instant benefit of single-photon counting is the elimination of readout noise. Time-gating and time-resolved detection also facilitate the reduction of noise from dark counts. The high time-resolution and high frame rate of SPAD imagers have been exploited to overcome sensitivity issues and demonstrate SPAD performance similar to that of electron multiplying charge-coupled device (EMCCD) and scientific CMOS cameras (Gyongy *et al.*, 2016b; Antolovic *et al.*, 2017).

When it comes to the application of SPAD technology in low-coherence interferometry and OCT, noise of the detector is not a major concern. The biggest challenges are detection efficiency in the NIR and saturation. Low detection efficiency of NIR photons is the primary reason for the relatively low sensitivity of the OCT system with the Ra sensors, compared to other technologies. For example, the photon detection probability (PDP) of Ra-II at 840 nm wavelength is 10 %, while the quantum efficiency (QE) of

the PointGrey camera (used in chapter 5) at this wavelength is 30 %. Furthermore, the fill factor (FF) of the pixels (15 % with the red SPADs of Ra-II, and 42 % with the PointGrey camera) also affects the amount of light collected with the sensor. SPADs with increased PDE at long wavelengths are subject to active research and sought for by other techniques as well, such as quantum communication (Zhang *et al.*, 2015), optical communication (Chitnis and Collins, 2014), diffuse optical tomography (Puszka *et al.*, 2013), and light detection and ranging (LIDAR) in particular (Takai *et al.*, 2016), therefore significant improvements are expected in this field. Detection efficiency can be tackled from two fronts, by increasing the PDP (which is essentially a similar measure to QE), or by increasing the FF (and therefore the active area) of the device. For the former, high detection probability, backside illumination (BSI) sensors and alternative SPAD structures and materials are suggested, such as InGaAs SPADs (Pellegrini *et al.*, 2006). Sensitivity loss due to low FF in traditional SPAD designs can be tackled with microlens arrays (Gyongy *et al.*, 2018). A novel technology, 3D stacking, addresses FF by separating the detection and processing of photons (Abbas *et al.*, 2016). Stacking of multiple tiers offers not only highly improved FF, but also allows novel SPAD structures to be efficiently combined with well established solutions for the surrounding circuitry of SPADs (e.g. combining massively parallel TDC structures with InGaAs SPADs). A similar approach also aims to detach detection from processing and decrease the complexity of the sensor, opposing to system on chip (SOC) solutions (Burri *et al.*, 2016). By moving counters and timing circuits to external processing units, such as reconfigurable FPGAs, the flexibility of the sensing platform increases as well. This is another example supporting the importance of complementary hardware in future sensor solutions.

The dilemma of focusing on weak signals (but saturating to strong ones) or adapting to strong signals (but not being sensitive to weak ones) is inherently present in any sensing technology. One of the main issues with SPAD sensors in OCT is to deal with strong optical reflections without saturating the detector. Even though the saturation level of SPAD based architectures is usually lower than that of CCD and CMOS cameras, the lower limit of detection, i.e. the noise floor, also proves to be lower in many cases. The ratio of these two, the dynamic range, is comparable in the case of many SPAD and CCD/CMOS sensors. The trade-off mentioned at the beginning of the paragraph can be eased by a high dynamic range sensor. One may think that the low saturation level of SPAD sensors is not necessarily a problem, provided that the sensor has low noise as well, and therefore a high dynamic range. If we assume, that the only noise source of the system is the noise of its sensor, than two different sensors with the same dynamic range has the same performance. Even though one of the sensors has a lower saturation level and lower noise, one would only need to attenuate the optical signals so that it falls within the range between these two. This is certainly not the

case. SPAD structures often have a high dynamic range (Eisele *et al.*, 2011), but the low sensitivity (low detection or quantum efficiency) results in weak detector output, which, in turn, have high uncertainties negatively affecting the achievable signal-to-noise ratio (SNR). This conclusion further emphasizes the need for high sensitivity (with regards to PDE) when working with SPAD sensors in reflectormetry and low-coherence interferometry. SPAD sensors need to have a higher saturation level as well. Short deadtimes and multiplexed pixels with short SPAD pulses are certainly beneficial, and alternative solutions have been investigated as well, such as 'XOR' summation of SPAD pulses (Gnecchi *et al.*, 2016).

When the saturating reflection origins from a confined region of the sample, the straightforward solution to the saturation problem is to use time-gating. Removal of saturating reflection by gating is only useful if SPADs can ignore the incident photons entirely. This requires gating by SPAD bias modulation. Unfortunately, the time needed for stable operation around the on/off switches of SPADs did not allow the gating technique to work at short distances between the region of interest and a strong reflection. The working principle of TCSPC, in theory, allows the recording of the depth profile even in the presence of strong reflectors that would saturate the sensor in single-photon counting (SPC) mode. This is because in TCSPC the detection of photons are well separated in time, from which the time profile of the signal is built through many exposure cycles. However, this assumes a low optical signal from the sample, which may need attenuation if being too strong. At an attenuation where the strong reflector complies with TCSPC requirements, the weakly scattering sites may not give enough signal to be detected and distinguished from noise. Furthermore, TCSPC operation is too slow for imaging non-fixed targets. Masking of photon counters is still possible, and (non-saturating) reflections could be removed from OCT images at distances with clinical relevance. The time-resolution of masking is still inferior to optical gating solutions (Muller and Fraser, 2009), on the other hand, gating the detected signals with time-resolved SPAD sensors offers higher flexibility at lower complexity. Improvements in the gating performance will eventually allow SPAD technology to improve low-coherence interferometry through time-gating.

Time-resolved detection of interferometric spectra is enabled by the histogram mode of the Ra-II sensor. The combination of time-resolved detection and interferometry opens up new and exciting techniques, for which time-resolved detection of SD OCT spectra was demonstrated first. Assigning the optical signals to time regions based on the photons' time of arrival, on one hand, requires high resolution timing circuits. The TDCs of Ra-II have a resolution of 50 ps, which relates to a depth of 7.5 mm. This depth can be potentially shortened with higher resolution timing circuits. Timing circuits used with SPADs typically have a resolution of a few tens of picoseconds (Bruschini *et al.*,

2019), however, resolution of a few picoseconds and even sub-picosecond is possible (Li *et al.*, 2016). On the other hand, it is equally important that timing uncertainties are low. The key message from the initial tests was that jitter limits the temporal sectioning capability, especially when a high SNR is to be maintained. When short temporal sections are to be taken from the recorded fringes, SNR is decreased by a limited amount of signal falling to the selected time bin. Furthermore, uncertainties and mismatch in the spectral timing channels introduces noise. Therefore, uniform timing across the pixels and short instrument response function (IRF) are required from the sensor for this technique.

Depth resolved detection of interferometric spectra is promising for extended depth SD OCT. Parallel acquisition of signals from multiple depths was reported using multiple beams, focused at slightly shifted depths (Holmes *et al.*, 2008; Standish *et al.*, 2010). For each beam, the collected signal is combined with its respective reference signal in one of multiple parallel channels. The interference signal of the channels are then turned to B-scans. The regions in focus on each B-scan are used to build a compound image, where the depth of focus is increased (Holmes and Hattersley, 2009). In these systems, signals of different depths are collected at the same time, but on parallel channels. The limitation of the system is the practical number of channels. In contrast, with a time-resolved line sensor a single channel can be used, and signals from different depths are separated in time. The limitation in this case is the depth separation, which is determined by the time resolution of the detector. Focusing the probe beam in each depth is still a requirement. One of the solutions could be the use of multiple beams, just as in the above mentioned system, but without the need of multiple fibre couplers and reference arm channels. In fact, one of the problems related with the multi-beam system, namely, a limited optical power per beam can be overcome by separating the excitation pulses in time, using sample arm fibres with slightly different lengths.

7.3 Outlook

The work presented in this thesis can be followed up in several directions. Straightforward continuation of the TRFS studies is to deploy the Ra-II sensor in spectral FLIM, and further improve spectral double kinetic experiments. A TRFS sensing platform with the Ra-II sensor and state-of-the-art FPGA (potentially with a DSP unit) is a good candidate for a compact, versatile and cost effective system for fluorescence correlation spectroscopy (FCS), with direct calculation of the autocorrelation function (Elson and Magde, 1974). Time-resolved Raman spectroscopy is another example where deploying the Ra-II sensor with its excellent performance is expected to deliver outstanding results. Low-coherence interferometry combined with time-resolved

photon counting can lead to many ideas where coarse distance measurements are to be fine-resolved (superresolution ranging), or where frequency-domain reflectometry is to be extended (extended depth OCT). One example application with SPAD sensors in low-coherence interferometry is time-gated OCT. Time-resolved detection with high accuracy could advance separation of single and multiple-scattering. These and other applications will be enabled by increasingly versatile CMOS SPAD arrays that are lower noise, faster and with higher PDP.

Appendix A

Data Acquisition Protocol in the TRFS Studies

This appendix describes details of the data acquisition protocol used in the time-resolved fluorescence spectroscopy (TRFS) studies of chapter 4.

Data was recorded using non-interruptible software calls provided by the Opal Kelly library to initiate data transfer from firmware to the host computer. The data acquisition function was iteratively called to get data into preallocated arrays in the computer's memory. The size of the allocated memory was determined by the datasize of a single pixel (16 + 27 bits of the photon counter and time-to-digital converter (TDC) of each pixel, padded to 64 bit as depicted in Fig. A.2c), the number of pixels per line (256 pixels · 64 bit data + 64 bit header = 2056 byte), the number of required lines per iteration and the number of iterations (Fig. A.1). In some cases the host computer could not read data from the first in first out (FIFO) buffer with sufficient speed causing a buffer overflow. Data from the sensor is lost until there is sufficient space in the buffer again, causing missing and incomplete lines (gaps, Fig. A.2a) in the arrays. Gaps in the data stream were identified by checking if the size of data between markers of the first pixels (Fig. A.2c) matches the size of a line. For any gaps the incomplete and corrupted lines (Fig. A.2b) were discarded during processing of the measurement data.

Accurate timing of the recorded lines was achieved using one of the scanning coordinates in the line header (see description of the firmware in section 4.2.2). Instead of scanning information, a counter of the firmware was connected to this coordinate increasing with each acquired line. To determine the timepoint of the recorded data with respect to the start of the experiment, the line counter value was used with the known timespan of a single line. The timespan of a single line is determined by the firmware's data readout method (see Fig. 4.7). With non-overlapping exposures and readouts the timespan of a line is the sum of the exposure time and approximately 47 µs of readout time. When the readouts are parallel to exposures there is still an idle cycle of 0.25 µs between subsequent exposures, due to firmware limitations. Similarly, the timespan

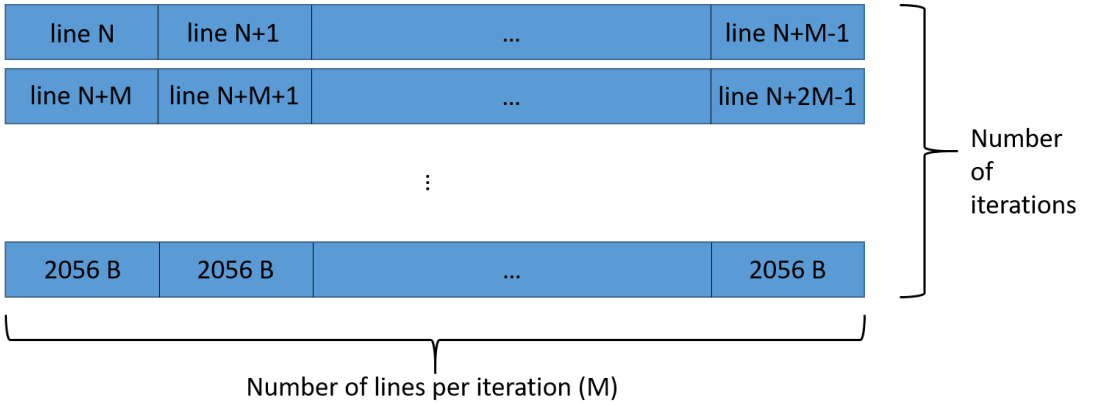

Figure A.1: The data from each read iteration is saved to rows of a two dimensional array. The size of a single row is determined by the number of requested lines per iteration and the size of one line (2056 byte).

of gaps was assessed based on the line counter value of the neighbouring complete lines and the number of lines between these with their known timespan.

During the data transfer in a single read iteration a few lines were lost introducing a gap with an average of 0.227 ms after every 248 lines from the sensor, which was a result of the FIFO buffer being filled-up. It follows from the above that the useful data with respect to the number of complete lines read from the sensor was somewhat lower than the number of requested lines. The number of lines lost during a single read iteration was less then 1 % of the number of requested lines in that iteration, throughout the whole experiment.

To capture the fast rise of chlorophyll fluorescence with no gaps between reading iterations, 3500 spectral lines were requested in a single iteration at the beginning of data acquisition (Fig. A.3a). Experimenting with larger array sizes showed that gaps of several milliseconds is introduced by the computer during an iteration when having to work with too many lines. Care had to be taken as well with respect to starting the laser illumination. If this has happened before recording the data, the beginning of the fast fluorescence rise would have been missed. Due to the limited size of the preallocated memory, the laser could not be started much later than initiating the data transfer either, as this would have lead to missing the end of the fast fluorescence rise. As a solution, the laser was started with a delayed software command right before calling the data reading function. 343 lines from the beginning of the recorded data array were omitted, belonging to the period during which the laser has not yet turned on.

In the second part of the experiment the acquisition of 1000 lines was initiated in 500 iterations and the recorded data stored in 500 rows of a matrix (Fig. A.3b). Due to

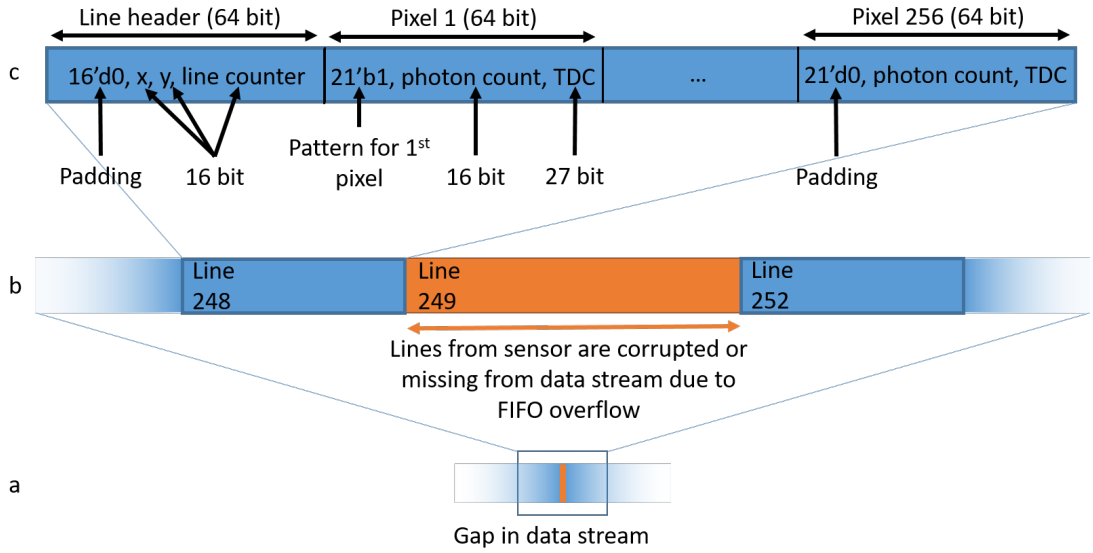

Figure A.2: Gaps are introduced in the data stream due to dropping of lines when the first in first out buffer overflows (a). The data stream has missing and corrupted lines in these cases, which are discarded during processing of the measurement data (b). The timespan of a gap is calculated using the line counter values in the line header (c) and the known acquisition time of a single line.

memory management on the host computer each data transfer was followed by an average gap of 6.4 ms. This gap was calculated using the line counter value of first line of an iteration and the last line of a previous one. In the third part of the protocol 500 times 1000 lines were acquired in a similar fashion but with an additional idle cycle in each iteration (Fig. A.3c). This was added by a software 'wait' call in order to reduce the data throughput and to accommodate to the more relaxed sampling requirements when measuring the slow intensity fall. The average gap between the recorded arrays was 99.1 ms.

Histograms of time-resolved fluorescence decays were built from the acquired time-correlated single-photon counting (TCSPC) lines. The number of TCSPC lines and number of pixels to bin together in the histograms were selected according to the sampling rate requirements, as described in chapter 4. Acquisition parameters such as exposure time, and number of lines recorded in a single block were slightly different in the TCSPC experiment with the Förster resonance energy transfer (FRET) probe, and lung auto-fluorescence. In the FRET probe fluorescence transient measurement, an exposure time of 8.3 µs was used followed by readout time. 1000 blocks of 1000 TCSPC lines were read from the sensor with no software delay and an additional 500 blocks of 1000 lines with an additional software delay of 100 ms. In the lung study, no change of the fluorescence lifetime was expected over time. 15 blocks of 5000 TCSPC lines were read with additional software delay of 0.5 s. In the chlorophyll study

using 'centre of mass' mode (CMM) the applied exposure time was 83.3 µs. 5000 lines were read initially, followed by 400 blocks of 1000 lines with no software delay and 300 blocks with 100 ms software delay. In each case, the gaps introduced by the host computer were similar as described previously. Processing steps of the CMM data is described in Appendix B.

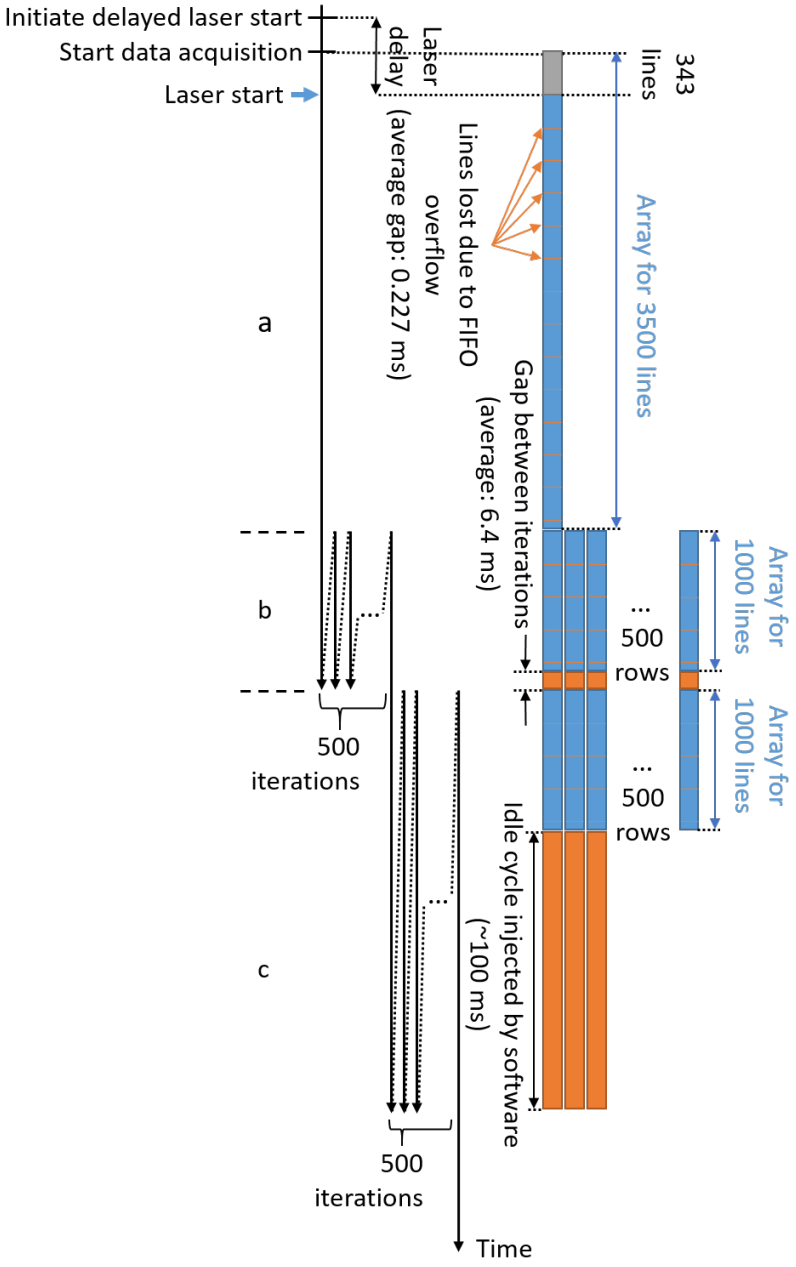

Figure A.3: Diagram of the data acquisition protocol used in the measurement of chlorophyll fluorescence. A single read iteration of 3500 lines (a) was followed by 500 iterations of 1000 lines with an average of 6.4 ms gaps between iterations (b), and another 500 iterations of 1000 lines with ~100 ms gaps between iterations (c). The timescale on the left is wrapped up around iterations for illustration purposes. The respective data arrays on the right does not depict the idle time of sensor readouts when non-overlapping data readout was used (refer to Fig. 4.7)

.

Appendix B

Correction of the Centre of Mass of Photon Arrivals for Fluorescence Lifetime Estimation

The 'centre of mass' mode (CMM) allows rapid lifetime estimation of fluorescence decays (Li *et al.*, 2010, 2012). In this mode, the time-to-digital converters (TDCs) of Ra-I are started by photon arrivals and stopped by the reference pulses of the source, but not restarted. At the end of an exposure cycle, two values are read out from the sensor for each pixel. The first is the summed arrival time, given by the TDC of the pixel as a digital code. The second value is the sum of the photons collected during the integration time (exposure time), similarly to single-photon counting (SPC). The centre of mass of photon arrivals at each pixel is calculated on the host computer as the fraction of the summed arrival time and the number of photons detected. In order for this value to make sense, further processing is required which is described in this appendix.

The centre of mass of photon arrivals τ_{cm} from a single CMM line can be mathematically written as

$$\tau_{CM}(p,T_i) = \begin{cases} \frac{\sum_{n=1}^{N(p,T_i)} \tau(n,p,T_i)}{N(p,T_i)} & : N(p,T_i) > 0 \\ 0 & : N(p,T_i) = 0 \end{cases} \tag{B.1}$$

where $\tau(n,p,T_i)$ is the arrival time of the n-th photon collected in the time interval of (integration time) T_i at pixel p, and $N(p,T_i)$ is the number of photons processed during the integration time T_i at pixel p. Here, the arrival time τ is actual time, which is calculated by multiplying the digital code of a TDC with the TDC resolution. In order to avoid division with zero, we define the centre of mass of photon arrivals as zero when no photons are detected on a pixel during the exposure. Often a threshold value is given instead as the number of detected counts below which τ_{CM} is defined as zero. This is to prevent highly inaccurate lifetime estimates due to low photon statistics. For

simplicity, the notation of the zero case of τ_{CM} is omitted hereafter.

In chapter 4, several CMM lines and several pixels along the lines were binned for more accurate estimates of the centre of mass. Let $T_{sum,i}$ denote the increased exposure interval when M CMM lines are added, that is the union of T_j integration intervals according to

$$T_{sum,i} = \bigcup_{j=(i-1)\cdot M}^{i\cdot M-1} T_j \tag{B.2}$$

The centre of mass of photon arrivals during the total interval of $T_{sum,i}$ at pixel p shall be written as $\tau_{CM}(p, T_{sum,i})$, which is

$$\tau_{CM}(p, T_{sum,i}) = \frac{\sum\limits_{T_j \in T_{sum,i}} \left(\sum\limits_{n=1}^{N(p,T_j)} \tau(n, p, T_j) \right)}{\sum\limits_{T_j \in T_{sum,i}} N(p, T_j)} \tag{B.3}$$

Let τ_{CM} now define the centre of mass of photon arrivals measured over several CMM lines (with a total integration interval of $T_{sum,i}$) and a set of several pixels (band B).

$$\tau_{CM}(B, T_{sum,i}) = \frac{\sum\limits_{p \in B} \left(\sum\limits_{T_j \in T_{sum,i}} \left(\sum\limits_{n=1}^{N(p,T_j)} \tau(n, p, T_j) \right) \right)}{\sum\limits_{p \in B} \left(\sum\limits_{T_j \in T_{sum,i}} N(p, T_j) \right)} \tag{B.4}$$

Note, that in the three cases above we used the same notation τ_{CM} for simplicity, even though the three definitions are different. It is easy to see from the above definitions that changing the delay of the sync signal (stopping the TDCs) increases the measured arrival times (τ) and shifts the centre of mass. Furthermore, the instrument response function (IRF) convolved with the true decay also shifts the centre of mass. This effect varies from pixel to pixel and requires correction, especially when the IRF width is considerable with respect to the decay lifetime. Previous studies suggested to perform a differential CMM to get a a lifetime estimate τ_{CMdiff}, where the centre of mass of the IRF (τ_{IRF}) is subtracted from that of the measured data (τ_{CM}) (Krstajić *et al.*, 2015; Poland *et al.*, 2016)

$$\tau_{CMdiff} = \tau_{CM} - \tau_{IRF} \tag{B.5}$$

If the IRF is measured involving the same optical path as a decay measurement from the sample (e.g. from scattering solution), the sync delay of the decay and IRF

measurements is the same, and therefore cancelled in the differential CMM. Even if the IRF and the decays are measured with a different delay, the same cancellation effect can be achieved by shifting the IRF histogram peak to the same bin as the decay peak (assuming that the time of the fluorophore excitation is negligible). In this work, cancellation of the delay was performed in a different way. The centre of mass values of both the decay and the IRF were first normalised to the histogram peak position, and then subtracted from each other (see equation B.6). In the time-resolved fluorescence spectroscopy (TRFS) studies of chapter 4, the centre of mass of the measured decay (τ_{CM}) was acquired in CMM. To get the peak of the decays (τ_{peak}), the sensor was switched to time-correlated single-photon counting (TCSPC) mode during the experiment, and spectral decays histograms were built. The IRF were measured in a separate set of measurements in TCSPC mode. For more accurate lifetime estimates an additional lookup table (LUT) can be used, however, this was not investigated further in this thesis.

$$\tau_{CMdiff} = (\tau_{peak} - \tau_{CM}) - (\tau_{IRFpeak} - \tau_{IRF}) \tag{B.6}$$

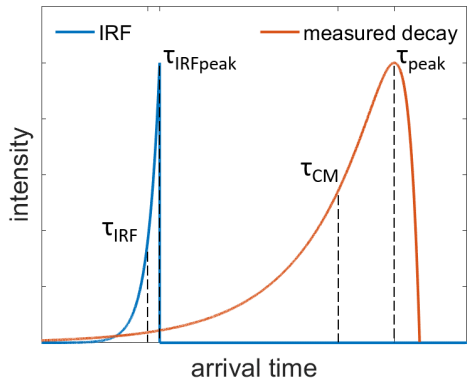

Figure B.1: Simulated example of a single exponential decay (red curve) convolved with the instrument response function (IRF) (blue curve) at a certain pixel. The time axis is reversed according to the 'stop-start' timing principle. The peak of the IRF ($\tau_{IRFpeak}$) and the measured decay (τ_{peak}), and the centre of mass of the IRF (τ_{IRF}) are retrieved from TCSPC data. These values are used for correction of the centre of mass of photon arrival times (τ_{CM}) given by the 'centre of mass' mode measurements.

Appendix C

Simulations

This appendix contains some of the Matlab (MathWorks) code used for the simulations in chapter 5, where various parameters of single-photon avalanche diode (SPAD) detectors were tested with their effect on the achievable count rate and signal-to-noise ratio (SNR). A major part of these studies was to simulate the output of a SPAD based pixel. The code for this is given in section C.1. In order to predict how the SNR of the optical coherence tomography (OCT) system is affected by SPAD parameters, interferometric spectra from a spectral-domain (SD) OCT setup were also simulated with respect to the number of photons falling to the sensor. These simulations were based on equations given by Drexler and Fujimoto (2015, eq. 2.6 in chapter 2). Simulations of interference spectra was also utilized in chapter 6, where the effect of mismatched gate positions across the spectrum was studied. Part of this code is shown in section C.2.

C.1 Simulations of a multiplexed SPAD pixel output

The following code was used to simulate the output of a pixel with several SPADs. In this simulation it is assumed that the SPAD output pulses are square shaped, and the 'OR' gate of the pixel is high when any of its inputs is high. The function takes arguments from the user such as the mean number or counts to generate ($N\_counts2gen$) and the dark count rate (DCR), the exposure time (T_exp), the deadtime of the SPADs and the length of shortened SPAD pulses ($T\_dead, Tpulse\_ns$). Each of the 8 SPADs of the simulated pixel can be enabled individually. Either a continuous source is simulated (CW) or a pulsed one, in which case the repetition rate of the source ($repRate$) and the jitter of pulses can be defined ($sigPulses$). Furthermore, the user can select if a uniform illumination model or a Gaussian beam is to be generated falling onto the pixel ($uniformIllum$). In the letter case the centre and width of the beam can be specified through parameters $muBeam$ and $Hbeam$.

```
function [N_incident, N_processed] = sim_paralyzablePixel(N_counts2gen,
    Texp_us,Tdead_ns,Tpulse_ns,muBeam,Hbeam,CW,repRate,sigPulses,plot_flag,
    SPADs_EN,DCR,uniformIllum)

% number of SPAD channels on detector (per pixel)
N_SPADs = 8;

% position resolution used for numeric calculations [um]
resH = 1e-4;
% SPAD pitch [um]
Hspad = 11.89;

% Gaussian beam profile across y axis (SPADs)
% sigma of beam, calculated from FWHM
sigBeam = Hbeam./2*sqrt(2*log(2));              % convert from FWHM to sigma

% axis along height
hBeam = 0:resH:N_SPADs*Hspad;                   % spatial positions along
    beam
beamPDF = gaussmf(hBeam,[sigBeam,muBeam]);      % define beam profile
beamPDF = beamPDF./sum(beamPDF);                % make beam profile to be a
    PDF (unity integral)

% array for probability of photon falling to SPAD
% i.e. percentage of element covering the beam
P_photon2SPAD = zeros(1,N_SPADs);
% coordinates belonging to element edges
for n_SPAD = 1: N_SPADs
    i_h_low = find(hBeam >= (n_SPAD-1)*Hspad , 1);
    i_h_high = find( hBeam > n_SPAD*Hspad , 1);
    if isempty(i_h_high)
        i_h_high = length(hBeam);
    end
    if uniformIllum
        % P_photon2SPAD(n_SPAD) = 1/N_SPADs;
        % NOTE, we test here the effect when the pixel consists
        % a certain number of SPADs
        P_photon2SPAD(n_SPAD) = 1/sum(SPADs_EN.' == '1');
    else
        P_photon2SPAD(n_SPAD) = sum(beamPDF(i_h_low:i_h_high-1));
    end
end

% NOTE the passed variable for the number of counts to generate is real
    valued
% in order to store array sizes we use a value instead of that that's
    rounded upwards
N_counts2gen = ceil(N_counts2gen);
```

```
% number of photon detections (per element)
N_countsDet = zeros(1,N_SPADs);

% start time end end time of photon detections per element
% NOTE here we overestimate the number of generated photons
% by assuming that all of them can fall to to a specific SPAD
riseTimes = zeros(N_SPADs,N_counts2gen);
fallTimes = zeros(size(riseTimes));

% start time and end time of photon detections on the pixel
riseTimes_joint = zeros(1,N_counts2gen);
fallTimes_joint = zeros(1,N_counts2gen);

% number of processed photons
N_countsProc = 0;

% exposure time [us]
Texp_ns = Texp_us*1e3;
if not(CW)
    % laser period [ns]
    Tlaser_ns = 1e3/repRate;

    % number of laser pulses in interval of exposure time
    N_laserPulses = floor(repRate*1e6*Texp_ns*1e-9);

    % number of counts to generate per laser pulse
    % is number of all counts to generate per number of laser pulses
    N_cnts2gen_perLaserPulse = N_counts2gen./N_laserPulses;
end

% number of incident photons
% NOTE this might be slightly different from the specified number of photons
% to generate, as statistical fluctuations (shot noise) are taken into
    account
N_incident = 0;

% SPAD enable
SPADisEN = ( bitand( bin2dec(SPADs_EN),pow2(0:N_SPADs-1) ) > 0) ;

% expected number of dark counts (array with values for each SPAD)
N_dark2gen = DCR.*1e-9*Texp_ns;

% ------------------------------------------------------------------
%               generate rising times of element output pulses
%
% NOTE continuous source will generate detection times with uniform
    distribution
```

```
% pulsed source will have a normal distribution of photons around a well
    defined timepoint
for n_SPAD = 1:N_SPADs

    % fprintf('generating counts on line %d\n',n_SPAD)

    % ----------  if SPAD is selected ----------------
    % perform any count generation only when the current SPAD is selected
    if SPADisEN(n_SPAD)

        % dark counts are uniformly distributed in time
        % assuming that dark counts have Poisson distributed values around
            measured DCR
        N_dark2gen_line = poissrnd(N_dark2gen(n_SPAD));

        % continuous source
        if CW
            % number of photons to generate to this element
            % NOTE shot noise is taken into account by generating a Poisson
                random number
            N_counts2gen_line = poissrnd(N_counts2gen * P_photon2SPAD(n_SPAD
                ));

            % uniform distribution of photon arrivals
            Trise = sort(0 + (Texp_ns-0).*rand(N_counts2gen_line +
                N_dark2gen_line,1));

            N_incident = N_incident + N_counts2gen_line;

        % pulsed source
        else
            % centre of laser pulses during the specified exposure time
            muT = (0:N_laserPulses-1)*Tlaser_ns;

            % number of photons per laser pulse to generate at this element
            % NOTE this is a mean value
            N_cntsPerLaserPulse_SPAD = N_cnts2gen_perLaserPulse *
                P_photon2SPAD(n_SPAD);

            % number of photons to generate at this element
            % is number of photons to generate at each pulse on this line
                times number of pulses
            % NOTE shot noise is taken into account by generating a poisson
                random number
            % anecdote
            % ideally each individual laser pulse carries photons that have
                a number with Poisson distribution,
            % however when summing Poisson distributed variables
```

```
        % (and multiplying the number of photons per pulse with the
            number of pulses is a summation)
        % the resulting value of the summed random variable will also
            have a Poisson distribution
        % with an expected value that's the sum of the individual
            expected values
        % hence we just saving some computation by only generating a
            Poisson random variable at this step
        N_counts2gen_line = poissrnd(N_cntsPerLaserPulse_SPAD *
            N_laserPulses);


        % anecdote
        % in order to generate a rising edge to all photons of an
            element
        % the centre of the laser pulse is used several times
        % by selecting multiple elements from the centre value
            population with replacement
        muT_line = randsample(muT,N_counts2gen_line,true);
        Trise_photons = normrnd(muT_line,sigPulses);

        % generate uniformly distributed dark counts
        Trise_dark = 0 + (Texp_ns-0).*rand(N_dark2gen_line,1);

        % consider dark noise and photons as well
        Trise = sort(horzcat(Trise_photons,Trise_dark'));

        % increase the number of incident photons
        N_incident = N_incident + N_counts2gen_line;

    end % if CW

    if length(Trise)
        % array to store (rising and falling) edges of SPAD pulses (no
            shortening)
        pulseEdges = zeros(2,length(Trise));
        % number of photons detected on this channel
        N_det = 0;

        % --------get rid of get rid of counts arriving during dead time
             ----------
        % -------------------- but make pulses longer
            ------------------------------
        % this is known by checking if the rising times are closer than
            the specified dead time
        % the first photon is surely detected, register in array
        pulseEdges(1,1) = Trise(1);
        if Trise(1)+Tdead_ns > Texp_ns
```

```
            pulseEdges(2,1) = Texp_ns;
        else
            pulseEdges(2,1) = Trise(1)+Tdead_ns;
        end
        N_det = 1;
        % iterate through all generated rising
        if length(Trise)>1
            for i_rise = 2:length(Trise)
                % if rise time is closer to the previous photon arrival
                    than dead time
                if Trise(i_rise) < (Trise(i_rise-1) + Tdead_ns)
                    % then dead time is prolonged according to the new
                        arrival
                    pulseEdges(2,N_det) = Trise(i_rise) + Tdead_ns;
                else
                    % otherwise the photon is detected with repsect both
                         to its rise and fall times
                    N_det = N_det + 1;
                    pulseEdges(1,N_det) = Trise(i_rise);
                    if (Trise(i_rise)+Tdead_ns) > Texp_ns
                        pulseEdges(2,N_det) = Texp_ns;
                    else
                        pulseEdges(2,N_det) = Trise(i_rise)+Tdead_ns;
                    end
                end
            end
        end
        % number of registered photons at this line
        N_countsDet(n_SPAD) = N_det;
        % rise times and fall times of registered photons
        riseTimes(n_SPAD,1:N_countsDet(n_SPAD)) = pulseEdges(1,1:
            N_countsDet(n_SPAD));
        fallTimes(n_SPAD,1:N_countsDet(n_SPAD)) = pulseEdges(2,1:
            N_countsDet(n_SPAD));

        % % do not go beyond exposure time with fall times
        % fallTimes(n_SPAD,fallTimes(n_SPAD,1:N_countsDet(n_SPAD))>
            Texp_ns) = Texp_ns;

        % --------get rid of get rid of counts arriving during dead time
             ----------

        % apply pulse shortening by setting fall times
        fallTimes(n_SPAD,1:N_countsDet(n_SPAD)) = riseTimes(n_SPAD,1:
            N_countsDet(n_SPAD)) + Tpulse_ns;


    end
```

```
    end % bitand(SPADs_EN,pow2(n_SPAD))
    % ----------  if SPAD is selected ----------------

end
%
%               generate rising times of SPAD pulses
% ------------------------------------------------------------------


% -------------------------------------------------
%                   sum up lines
%
% take care of overlapping photon detections
% iterate through all SPADs/lines
for n_SPAD = 1:N_SPADs

    % ----------  if SPAD is selected ----------------
    % perform any count generation only when the current SPAD is selected
    if SPADisEN(n_SPAD)

        % if there's no photon detected yet
        if N_countsProc == 0
            riseTimes_joint(1:N_countsDet(n_SPAD)) = riseTimes(n_SPAD,1:
                N_countsDet(n_SPAD));
            fallTimes_joint(1:N_countsDet(n_SPAD)) = fallTimes(n_SPAD,1:
                N_countsDet(n_SPAD));
            % number of processed photons
            N_countsProc = N_countsDet(n_SPAD);
        else    % some photons are already detected
            % for each photon pulse of the current SPAD
            for n_pulse = 1:N_countsDet(n_SPAD)

                % position of the pulse edges wrt rising and falling edges
                    of the combined line
                risingBeforeRiseTimes = riseTimes(n_SPAD,n_pulse) <
                    riseTimes_joint(1:N_countsProc);
                risingBeforeFallTimes = riseTimes(n_SPAD,n_pulse) <
                    fallTimes_joint(1:N_countsProc);
                fallingBeforeRiseTimes = fallTimes(n_SPAD,n_pulse) <
                    riseTimes_joint(1:N_countsProc);
                fallingBeforeFallTimes = fallTimes(n_SPAD,n_pulse) <
                    fallTimes_joint(1:N_countsProc);

                % index of the rising and falling edges of the SPAD pulse
                % wrt to rising and falling edges of the combined line
                idx_rise2rise = find(risingBeforeRiseTimes,1);
                idx_rise2fall = find(risingBeforeFallTimes,1);
                idx_fall2rise = find(fallingBeforeRiseTimes,1);
```

```
idx_fall2fall = find(fallingBeforeFallTimes,1);

% in case edge is above all edges of the combined line
if(isempty(idx_rise2rise))
    idx_rise2rise = N_countsProc+1;
end
if(isempty(idx_rise2fall))
    idx_rise2fall = N_countsProc+1;
end
if(isempty(idx_fall2rise))
    idx_fall2rise = N_countsProc+1;
end
if(isempty(idx_fall2fall))
    idx_fall2fall = N_countsProc+1;
end

% 1) rising of pulse is between a rising and falling edge on
     the combined line
%                                 __________
%       combined line:  ______|          |__________
%                                  ___
%       pulse:          __________|    .....


if  idx_rise2rise > idx_rise2fall

%   1A)
%       falling of pulse is between the same rising and
    falling edges
%                                 __________
%       combined line:  ______|          |__________
%                                   _____
%       pulse:          ________|     |_____________


%   if idx_rise2fall == idx_fall2fall

% %         the pulse has to be ignored in this case

%   end

%   1B)
%       falling of pulse is after the falling edge, but
    before the next rising edge
%                                 __________
%       combined line:  ______|          |__________
%                                      _____
%       pulse:          _____________|     |_____________
```

```
    if idx_fall2rise == idx_fall2fall

%       fall time of respective combined line pulse has to
    be extended
        fallTimes_joint(idx_rise2fall) = fallTimes(n_SPAD,
            n_pulse);

    end

%   1C)
%       falling of pulse is after the falling edge, and
    after the next rising edge
%                                 __________     ______
%       combined line:  ______|          |___|      |_______
%                                      ________
%       pulse:          _______________|        |______

    if idx_fall2rise > idx_rise2rise

%       the two pulses of the combined line have to be
    merged to one
        fallTimes_joint(idx_rise2fall) = Inf;
        riseTimes_joint(idx_rise2rise) = Inf;
        riseTimes_joint(1:N_countsProc) = sort(
            riseTimes_joint(1:N_countsProc));
        fallTimes_joint(1:N_countsProc) = sort(
            fallTimes_joint(1:N_countsProc));
        N_countsProc = N_countsProc - 1;
    end

% 2) rising of pulse is before a rising on the combined line
%                                 __________
%       combined line:  __________|          |__________
%                            ___
%       pulse:          ___|    .....

elseif  idx_rise2rise == idx_rise2fall

%   2A)
%       falling of pulse also happens before the rising edge
     on the combined line
%                                 __________
%       combined line:  __________|          |______
%                            ___
%       pulse:          ___|   |____________________

    if idx_rise2rise == idx_fall2rise
```

```
            %       the pulse has to be registered on the combined line
                    N_countsProc = N_countsProc + 1;
                    riseTimes_joint(N_countsProc) = riseTimes(n_SPAD,
                        n_pulse);
                    fallTimes_joint(N_countsProc) = fallTimes(n_SPAD,
                        n_pulse);
                    riseTimes_joint(1:N_countsProc) = sort(
                        riseTimes_joint(1:N_countsProc));
                    fallTimes_joint(1:N_countsProc) = sort(
                        fallTimes_joint(1:N_countsProc));
                end

            %   2B)
            %       falling of pulse happens after the rising edge but
                before the falling edge on the combined line
            %                                  __________
            %       combined line:  __________|          |______
            %                                ___
            %       pulse:          ________|   |____________________

                if (idx_fall2rise > idx_fall2fall)

            %       rising edge of pulse on the combined line has to be
                aligned to rising edge of pulse
                    riseTimes_joint(idx_rise2rise) = riseTimes(n_SPAD,
                        n_pulse);
                end

            %   2C)
            %       falling of pulse also happens after the rising and
                falling edges on the combined line
            %                                  __________
            %       combined line:  __________|          |______
            %                                ________________
            %       pulse:          ________|              |____

                    % NOTE, this shouldn't happen anyway


            %  3) pulse is after all other pulses of the combined line
            % elseif

            end % cases of different SPAD pulse position

        end % iterating through each photon pulse of the current SPAD

    end % performing addition of pulses
```

```
    end % ----------  if SPAD is selected ----------------

end % iterating through SPADs

% number of processed photons (after ORing SPADs)
N_processed = N_countsProc;
```

C.2 Simulations of interference spectra in an SD OCT setup

The following code is part of the simulations for generating an interference spectrum from a spectral-domain optical coherence tomography (SD OCT) setup. The user can specify the wavelength range ($lambda\_low\_sim$ and $lambda\_high\_sim$) of the light source with Gaussian spectrum, the power entering the interferometer (P_spect), the attenuation at the sample and reference arms (att_R_dB,att_S_dB) and the path difference (z_diff), and various sensor parameters such as fill factor (FF) and photon detection efficiency (PDE).

```
% integration time [us]
Texp_us = 100;
% fill factor (approx.)
FF = 0.4;
% photon detection efficiency (approx...)
PDE = 0.1;

% wavelength range
% low wavelength value [nm]
lambda_low_sim = 823;
% high wavelength value [nm]
lambda_high_sim = 887;
% centre wavelength [m]
lambda0 = 850e-9;
%
% power entering interferometer [W]
% NOTE assuming optical setup with 1 efficiency
P_spect = 0.8e-4;
%
% reflectors, attenuation in arms and spectrometer
% NOTE assume single reflector in sample arm
% path difference compared to reference reflector
z_diff = 100e-6;
% distance of reference reflector [m]
zR = 15e-2;
% attenuation at reference arm
```

```
att_R_dB = 1;
% attenuation at sample arm
att_S_dB = 4;
%
% spectrometer efficiency (based on previous measurements)
nu_spect = 0.8;
%
% ------ consts ---------
% % speed of light in vacuum [m/s]
c_mps = 3e8;
% Planck const.
h = 6.6e-34;

% ----------------------
% simulate single reflector generated fringes on top of gaussian laser
    spectrum

% simulation of sinusoidal fringes is done is frequency space, we then move
    to wavelength space

% step size  [nm]
lambda_step_sim = (lambda_high_sim-lambda_low_sim)/(Npix-1);
% wavelength array
lambda_arr = (lambda_low_sim:lambda_step_sim:lambda_high_sim).*1e-9;
% wavelength array - nm
lambda_arr_nm = lambda_arr.*1e9;

% wave number array - spatial frequency
% k = 2pi/lambda
k = 2*pi./(lambda_arr);

% mean value of gauss shaped source electric field amplitude in wavenumber
    [1/m]
source_ampl_mean_k = (k(1)+k(end))/2;
% std deviation of gauss shaped source electric field amplitude in
    wavenumber [1/m]
source_ampl_sig_k = (k(1)-k(end))/6;

% energy of photon [J]
Eph = h*c_mps/lambda0;
% exposure time [s]
Texp_s = Texp_us.*1e-6;

% dead time (of a single SPAD) [ns]
Tdead = 36;

% number of photons expected
N_ph = P_spect*Texp_s/Eph*FF*PDE*nu_spect;
```

```
%
S_k = N_ph .* (1/(sqrt(2*source_ampl_sig_k^2*pi))).*gaussmf(k,[
    source_ampl_sig_k,source_ampl_mean_k]);

% distance of sample reflector
zS0 = zR-z_diff;

% reflectors, attenuation in arms and spectrometer
%
% electric field reflectivity of reference reflector
% NOTE assume non-complex reflection of 1 (plus attenuation)
rR = 1/(10^(2*att_R_dB/10));
%
% electric field reflectivity of sample reflector
% NOTE assume non-complex reflection of 1 (plus attenuation)
rS0 = 1/(10^(2*att_S_dB/10));
%
% power reflectivity is magnitude squared of electric field reflectivity
% power reflectivity of reference reflector
RR = (abs(rR))^2;
% power reflectivity of sample reflector
RS0 = (abs(rS0))^2;
%

% detected signal on spect
% NOTE no autocorrelation due to a single reflector in sample arm
ID_k = 1/4*(S_k*(RR+RS0))...
    + 1/4*(S_k.*(sqrt(RR*RS0)).*(exp(1i*2*k*(zR-zS0))+exp(-1i*2*k*(zR-zS0)))
        );
```

Appendix D

OCT Processing

In this appendix a minimal example script is shown used for processing the measured interference spectra, and generate OCT images (B-scans). The script was composed from several commands of the Matlab scripts used throughout the OCT studies.

```
% number of pixels on the sensor
Npix = 1024;

% ...

% settings for FFT
% wavelengths on the edges of the detected spectrum [nm]
lambdaLow = 823;
lambdaHigh = 887;

% number of sampling points along line
noPointsLine = Npix;
% set this value to a power of 2 higher than then
% number of pixels is zero padding is to be carried out
fftSize = Npix;

% wavelength vector [m]
lambdaArray = linspace(lambdaLow,lambdaHigh,noPointsLine).*1e-9;
% wavenumber vector [1/m]
kArray = 1./lambdaArray;
% wavenumber vector, evenly spaced
kArrayEvenlySpaced = linspace(1/lambdaHigh,1/lambdaLow,noPointsLine);

% ...

% subtract background
line_corrected = interference_spectrum - bg_spectrum;
% interpolate line to evenly spaced k array
kSpectrum = interp1(kArray,vals,kArrayEvenlySpaced,'spline');
% filter line to suppress the sides
line_filtered = line_interf.*hann(noPointsLine)';
```

```
% sampling interval
dk = (1/lambdaLow-1/lambdaHigh)/(fftSize-1);
AscanX = -1/(2*dk):1/(dk*fftSize):1/(2*dk);
% x axis of the A-scan (lengths)
AscanX = 0.5.*AscanX;

% FFT the spectrum
AscanY = abs(fft(line_filtered,fftSize)).^2;

% compress by taking loagrithm, and store in Bscan array
Bscan(:,n_line) = 10*log10(AscanY(1:fftSize/2));
```

Appendix E

Selection of Optical Coherence Tomography Images

E.1 Ra-I

Several images of a fingertip are shown in Fig. E.1, taken with the pulsed light source and the Ra-I sensor with blue SPADs enabled. The B-scans are of relatively low quality due to the sensitivity of the sensor.

Repeated structures often appear in the OCT B-scans when the sinusoidal modulations of the recorded interference spectrum are distorted. The repeated structures are a direct consequence of harmonics appearing on the A-scans. Figure E.2 shows several cases where strong reflections from the top layer of a fingertip caused harmonics with the Ra-I sensor.

E.2 Ra-II

OCT images of a fingertip taken with the Ra-II sensor are shown in Fig. E.3- E.6 when different types of SPADs were activated and different light sources were applied. No difference can be seen in image quality when cross-comparing the two types of SPADs. Vertical white lines in some of the images are caused by occasional reading of faulty frames from the sensor, caused by a bug in the communication layer of the Opal Kelly library. In a few images, a faint horizontal line is also visible close to zero depth. This line was observed to appear temporarily during imaging and is believed to be caused by the light source.

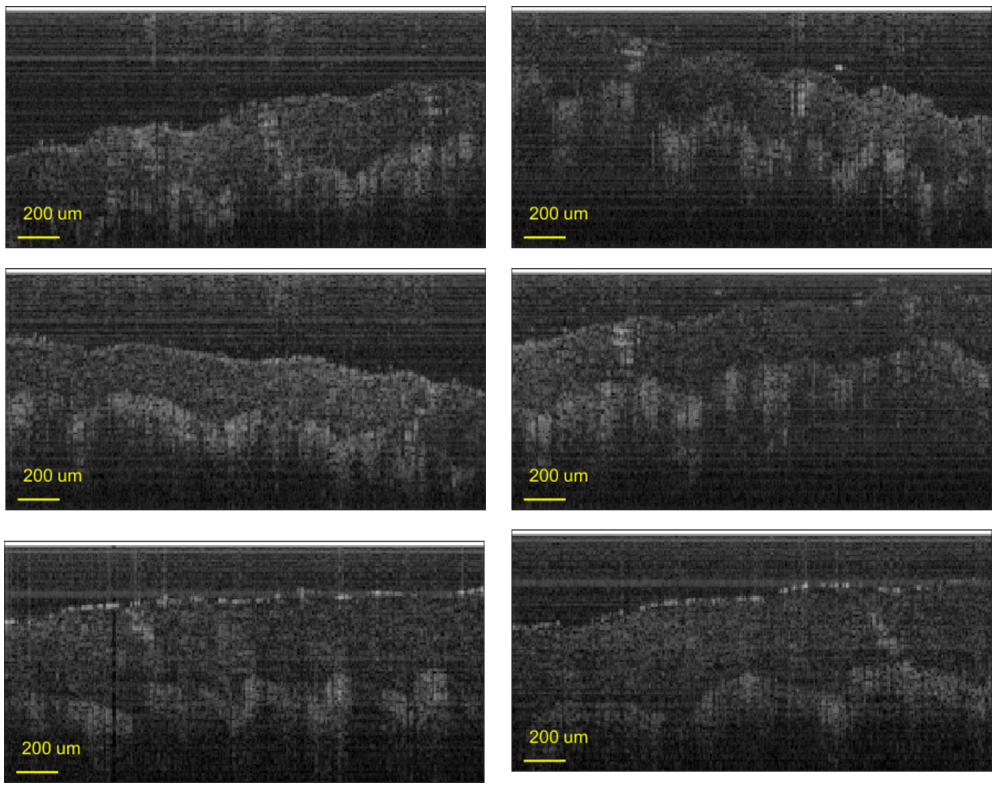

Figure E.1: OCT images of a fingertip taken with the Ra-I sensor, using the blue SPADs. Ten lines of the interference spectrum were recorded, and their respective A-scans averaged, resulting in an overall exposure time of 2 ms per scanning position, and 0.6 s per B-scan.

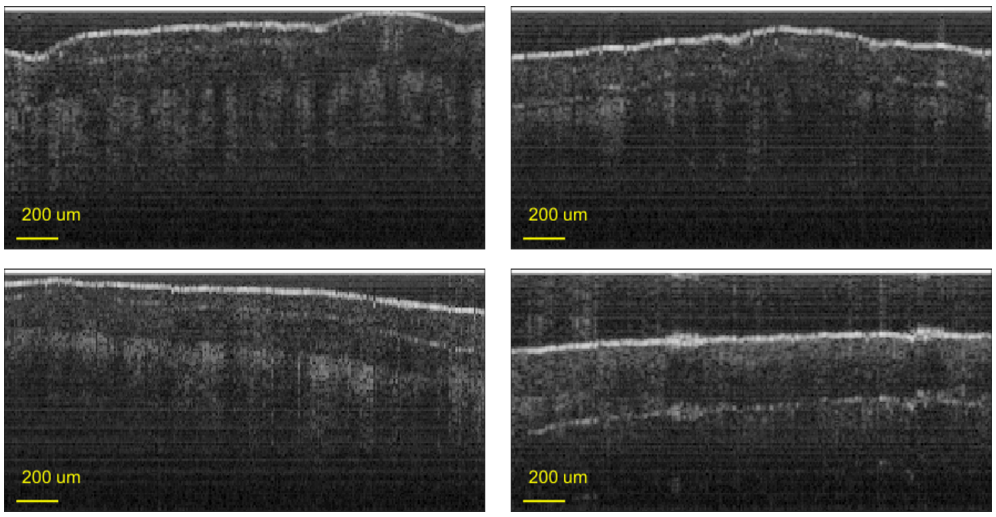

Figure E.2: OCT B-scans where the top layer of a fingertip are repeated, due to strong optical reflections in the detector's saturation region. The applied sensor was the Ra-I sensor with blue SPADs activated. The exposure time at each scanning position was 2 ms (averaged from 10 A-scans of 200 µs).

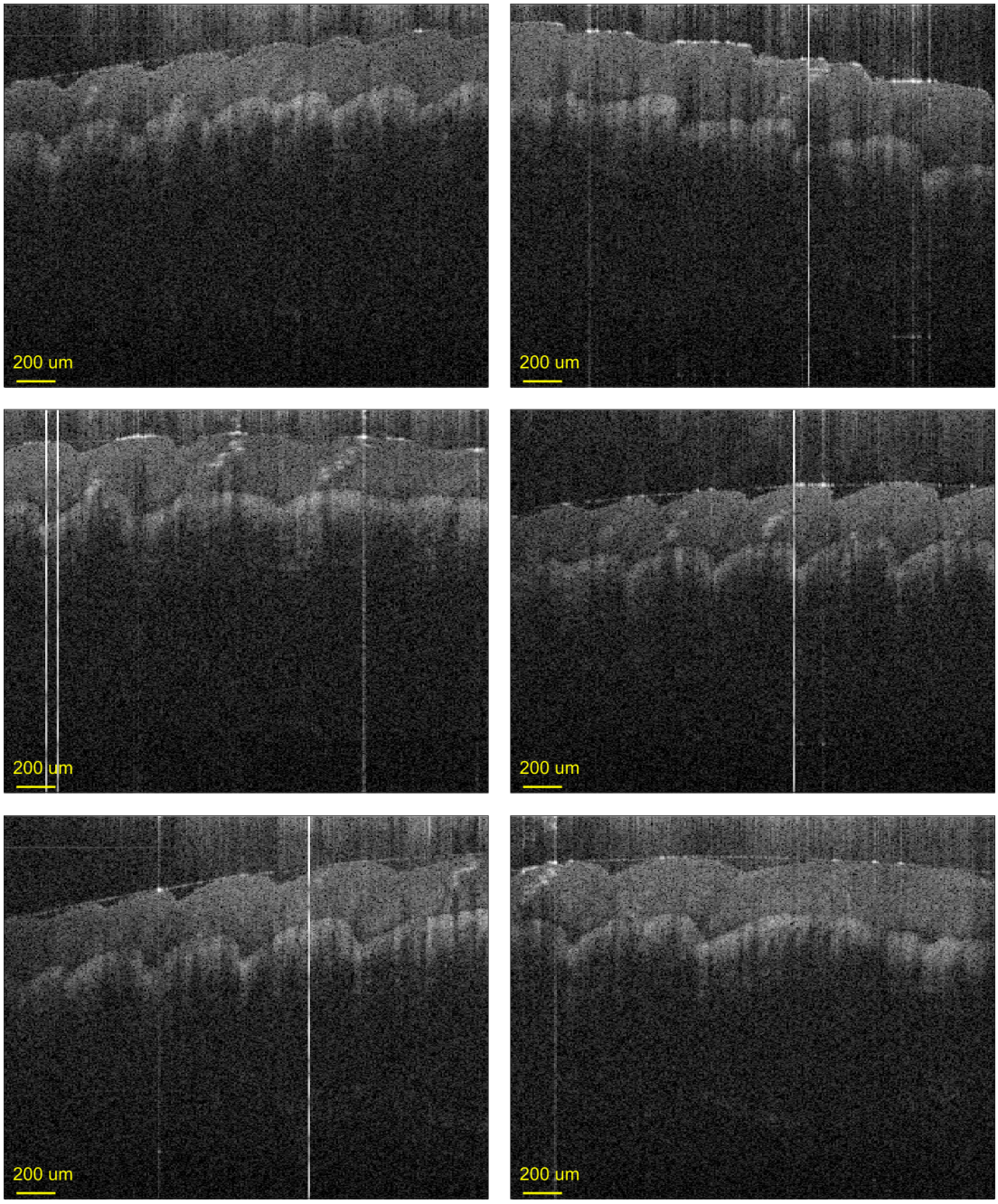

Figure E.3: OCT images of a fingertip taken with the Ra-II sensor and the Exalos superluminescent diode. The sensor was configured to use 1024 pixels along the spectrum. The applied SPADs were the blue SPADs. The applied exposure time was 1 ms per A-scan.

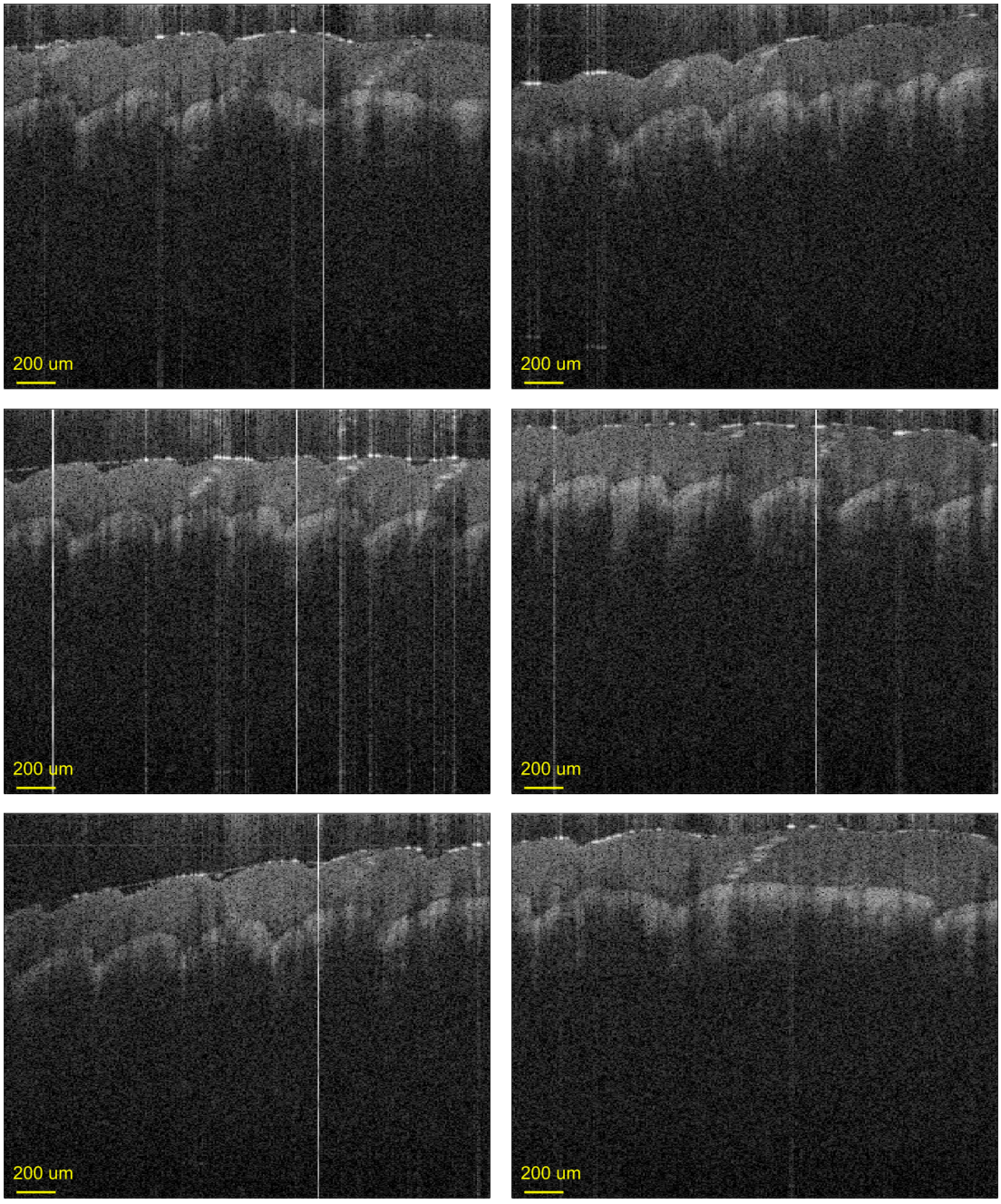

Figure E.4: OCT fingertip images captured with Ra-II, using all 1024 pixels and the red SPADs. The applied light source was the Exalos superluminescent diode and the exposure time was 1 ms per A-scan.

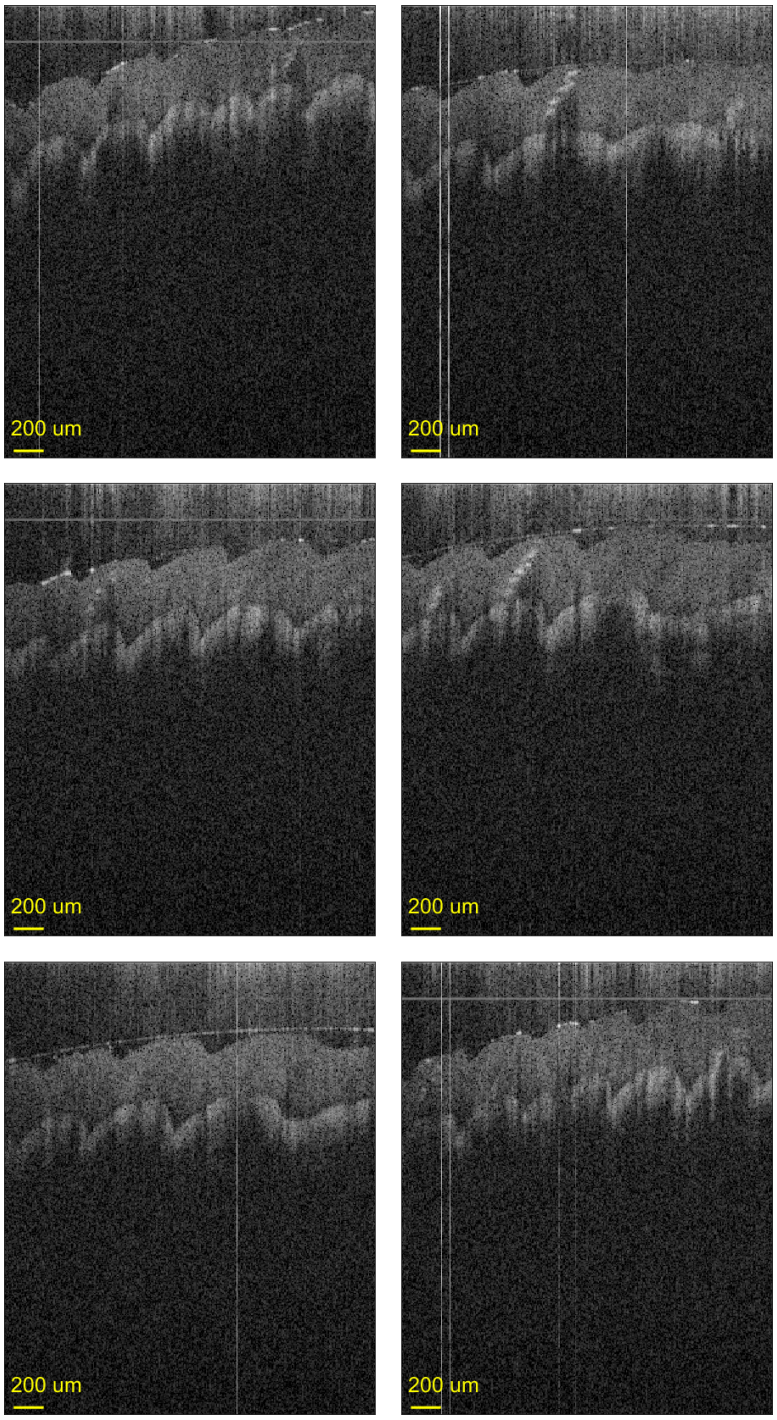

Figure E.5: OCT images of a fingertip using 1024 blue pixels of Ra-II. The applied light source was the filtered supercontinuum laser, and the applied exposure time was 1 ms per A-scan. The depth of the images is different compared to those with the Exalos superluminescent diode, due to the slightly different spectral bandwidth of the two sources.

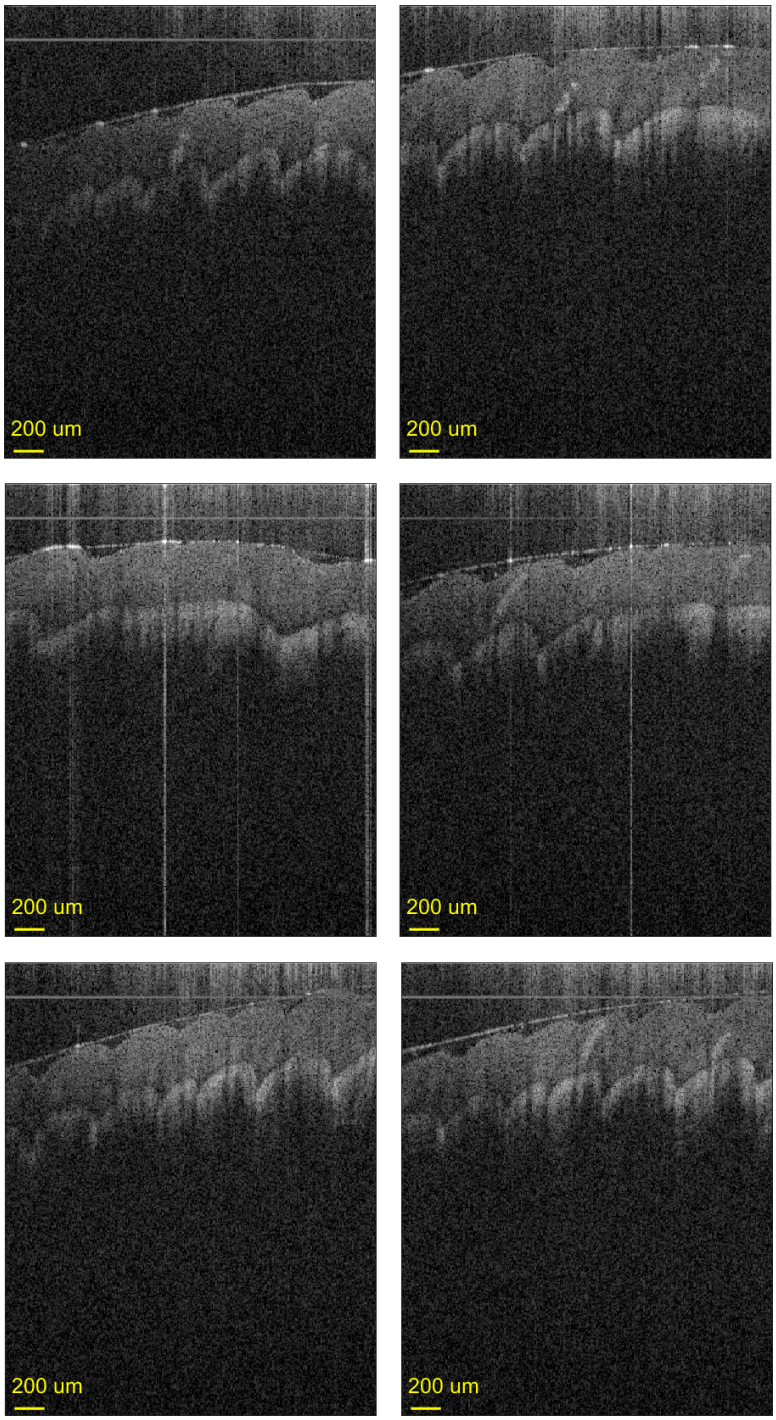

Figure E.6: OCT images of a fingertip using the Fianium light source, and red SPADs of Ra-II. The applied exposure time was 1 ms per A-scan.

E.3 Point Grey CMOS camera

In comparison with the Ra sensors, OCT images of a fingertip were also taken with a commercial complementary metal-oxide-semiconductor (CMOS) camera (Point Grey) with high efficiency in the near infrared (NIR) range. Figure E.7 shows images with 1 ms exposure time, and Fig. E.8 shows images with 0.4 ms exposure time per A-scan. The Point Grey camera has a higher quantum efficiency (QE) than the Ra sensors, and the OCT images taken with this camera have more details along the depth, accordingly. The images taken with the Point Grey camera have a lower transversal scanning range, in order to accommodate the low frame rate of the camera. A-scans were generated at each scanning position by summing 10 lines of 2048 pixels of the acquired frames. Due to the number of pixels per line (2048), the sampling rate of the interference spectrum was higher than that with the Ra-sensors. The depth of the OCT images is therefore higher. Still, due to high scattering of the applied light in tissue, the information (i.e. visibility of structures) is still limited. CMOS and charge-coupled device (CCD) cameras have a limited well depth for which shorter exposure times are preferable, however no saturation was observed either with 0.4 ms or 1 ms exposure times.

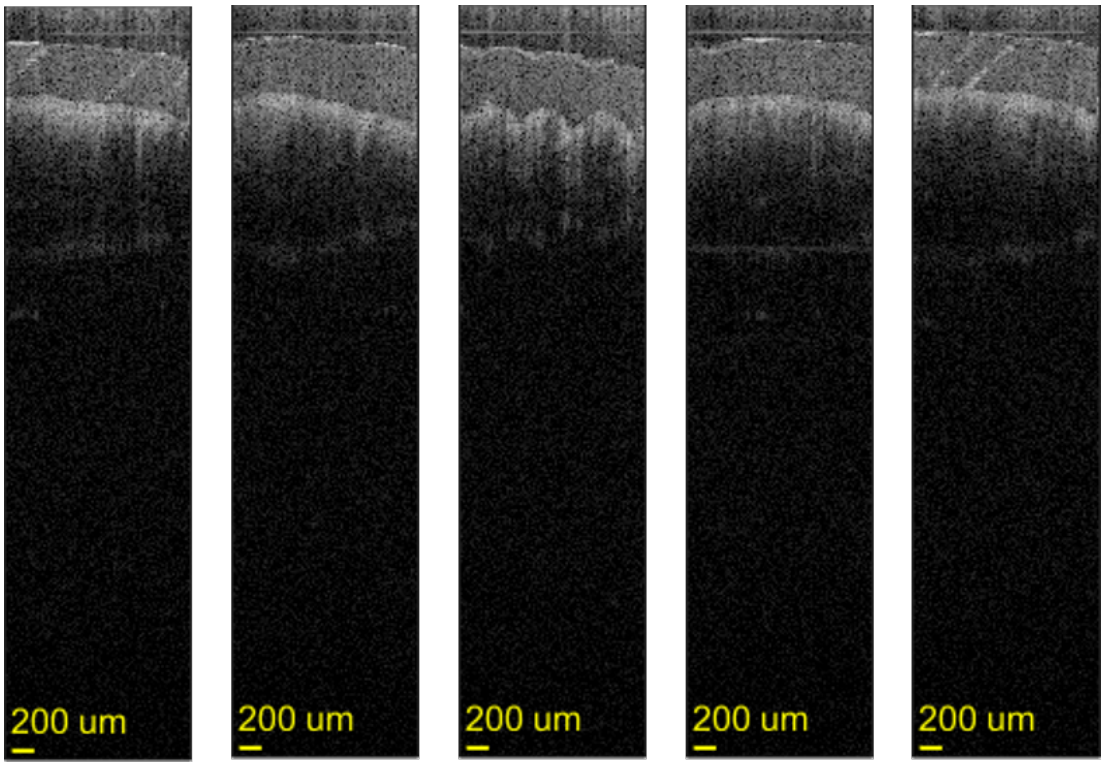

Figure E.7: OCT images of a fingertip taken with the Point Grey camera and an exposure time of 1 ms per A-scan.

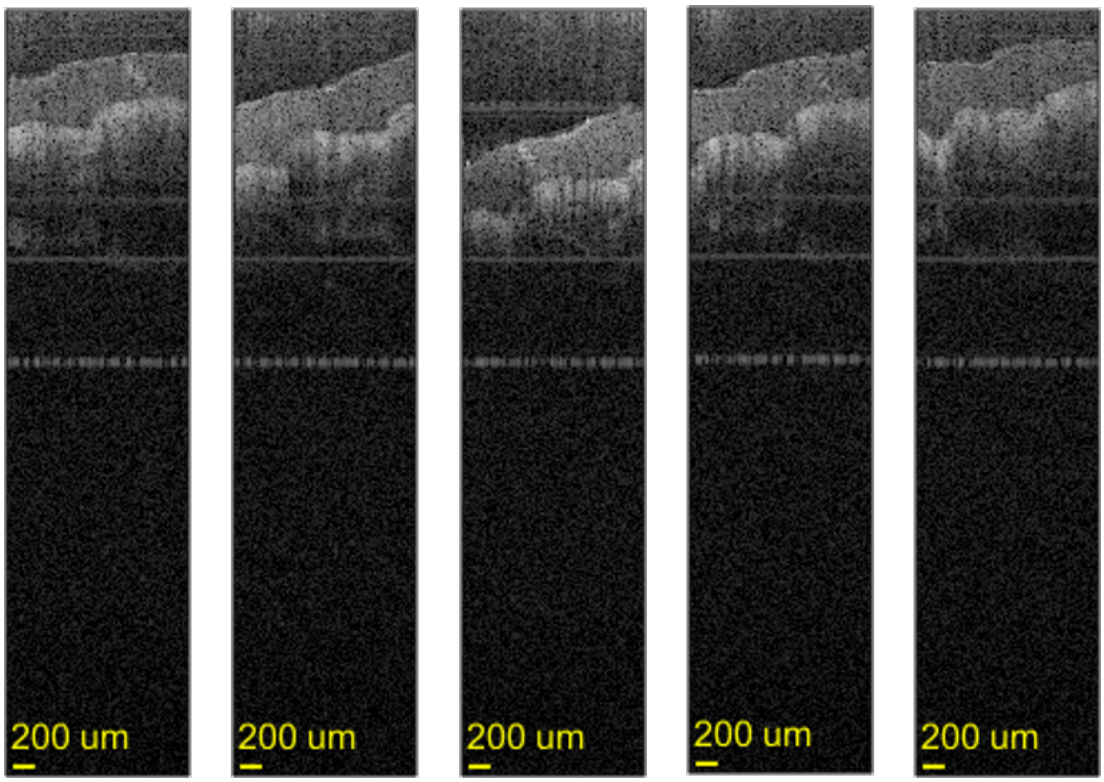

Figure E.8: OCT images of a fingertip taken with the Point Grey camera and an exposure time of 0.4 ms per A-scan.

E.4 Time-gated OCT images

Experiments were carried out for the removal of features along the optical path of an OCT setup, using the Ra sensors. For this, a glass slide was placed along the sample arm, close to the sample (fingertip). Reflections from the glass slide were removed from the images by masking the photon counters of the Ra sensors for a user defined time interval. Figure E.9 shows pairs of OCT images with and without temporal masking, using the Ra-I sensor. The distance between the fingertip and the glass slide was 10 cm. The applied exposure time was 200 µs per A-scan. In Fig. E.10, image pairs of a similar arrangement are shown, where the applied sensor was Ra-II with blue SPADs, and the exposure time was 1 ms per A-scan. The distance between the glass slide and the fingertip was 3 cm.

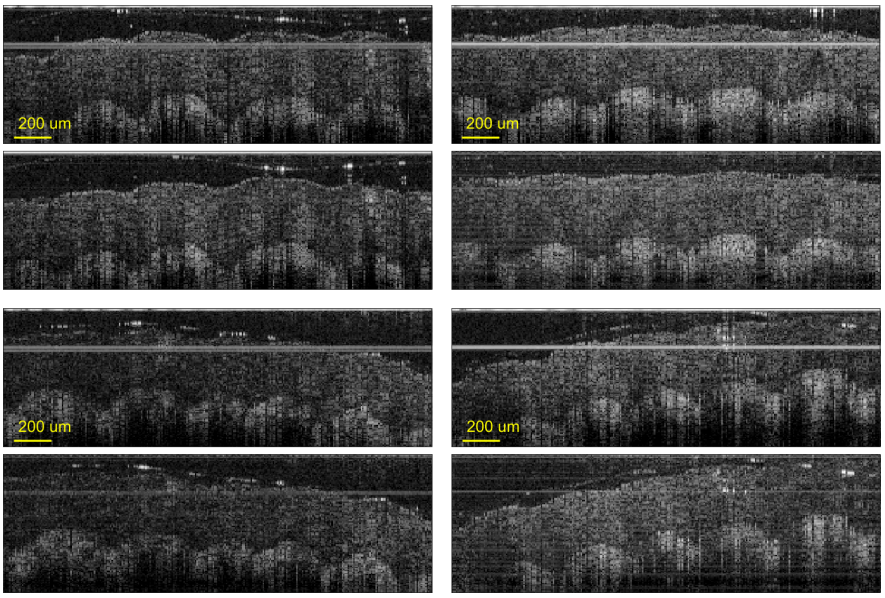

Figure E.9: OCT image pairs of a fingertip and a glass slide without and with masking of Ra-I. The distance between the glass slide and the fingertip was 10 cm, and the applied exposure time was 200 µs at each transversal scanning point.

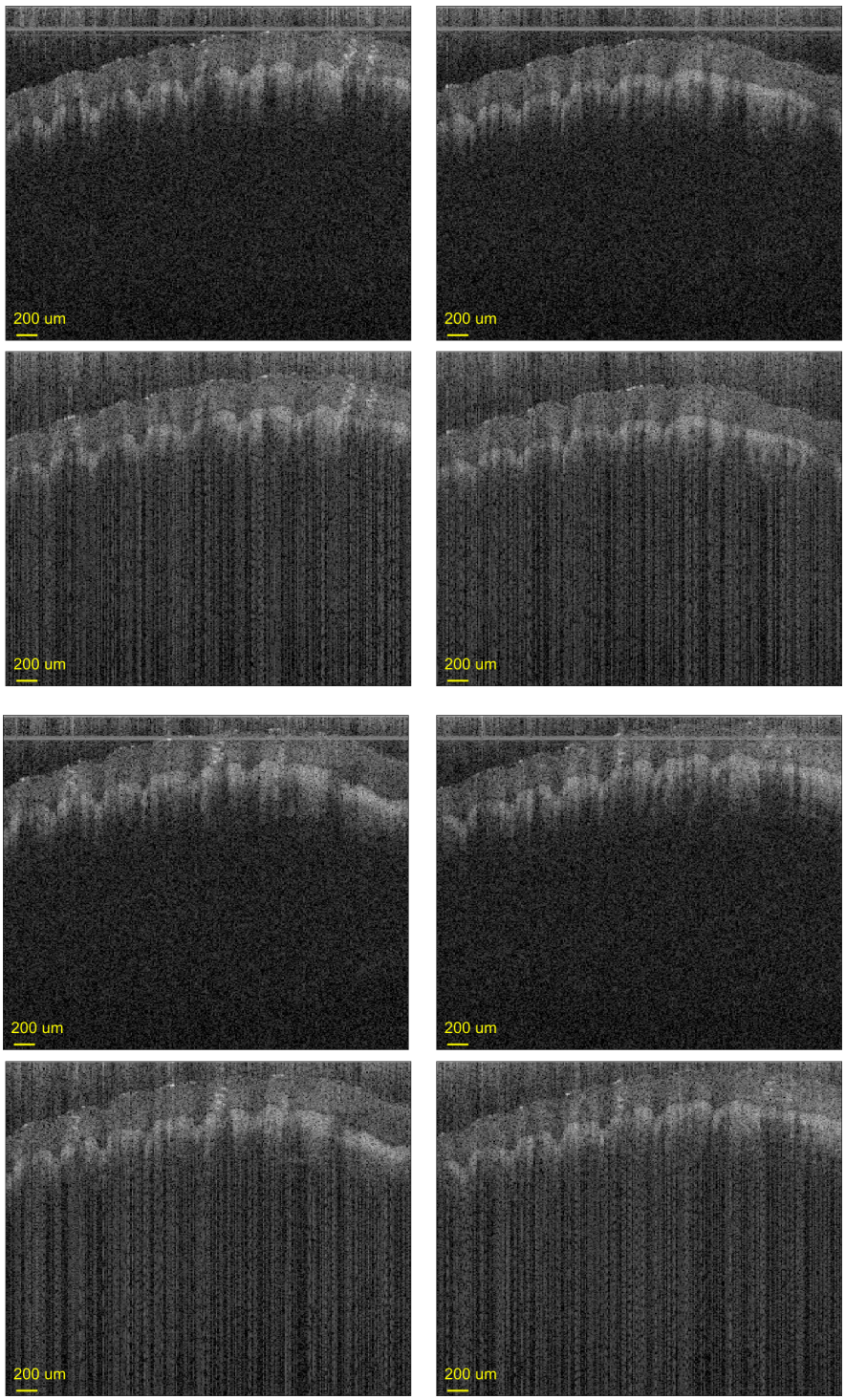

Figure E.10: OCT image pairs of a fingertip and a glass slide at 3 cm distance, without and with temporal masking applied. The images were taken with the Ra-II sensor with an exposure time of 1 ms at each scanning point.

E.5 Time-resolved OCT images

OCT images of a fingertip were also captured in histogram mode of the Ra-II sensor. In this mode, a spectral time-resolved histogram was collected at each scanning point. The interferometric spectral lines were generated by summing 8 time-bins of the time-resolved histograms, each with ~50 ps resolution. The interference spectra were processed according to standard SD OCT techniques, described earlier. The applied exposure time was 1 ms. In Fig. E.11, a fingertip was used only as a sample to show how OCT images can be retrieved from the spectrally and temporally resolved histograms. In a separate experiment, a glass slide was placed along the optical beam, at 4 cm distance of the fingertip. The glass reflection is removed from the images recorded in histogram mode, shown in Fig. E.12, by only selecting time bins containing the optical signal from the fingertip.

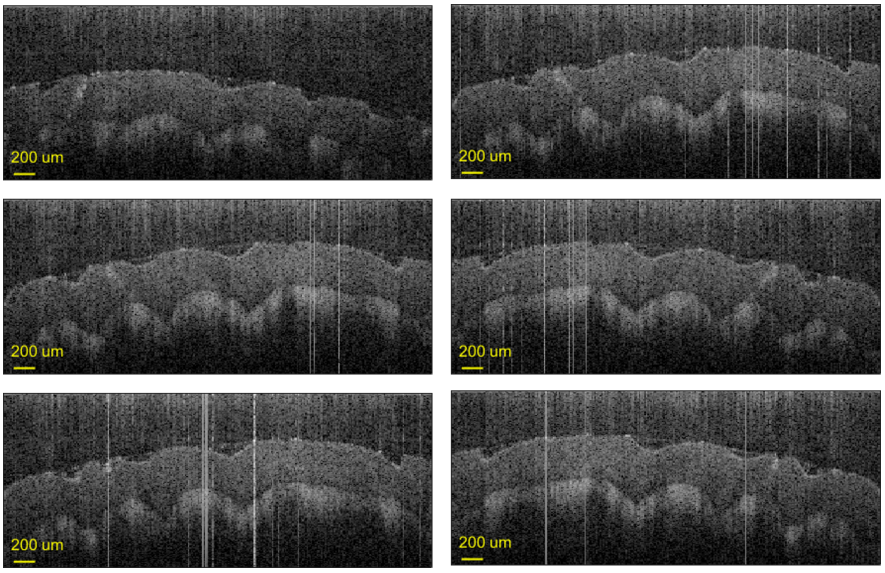

Figure E.11: OCT images of a fingertip recorded in histogram mode of the Ra-II sensor. While the demonstrated images contain only the optical signal backscattered from the fingertip, the technique allows sectioning along the sample depth based on the time-of-arrival of optical signals. The applied exposure time was 1 ms.

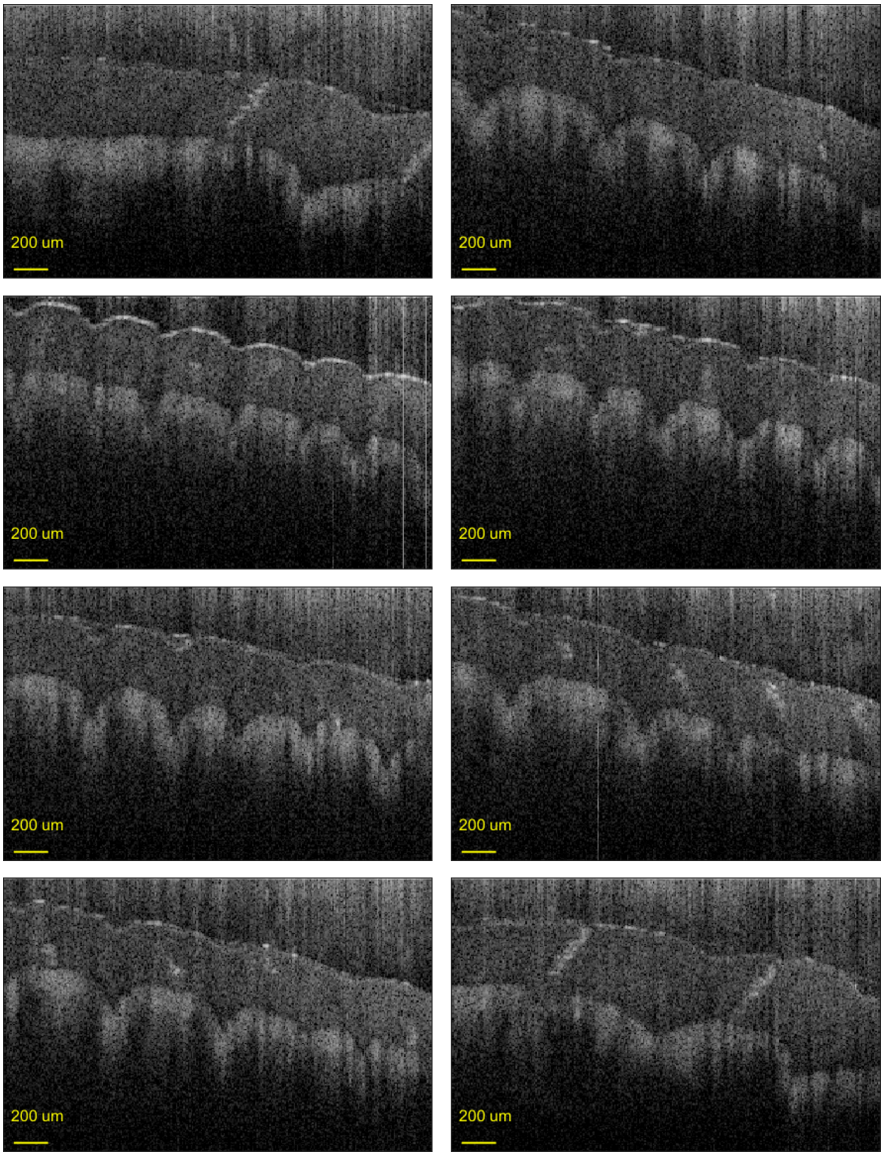

Figure E.12: OCT images of a fingertip and a glass slide at 4 cm distance, taken in histogram mode of the Ra-II sensor. Using temporal sectioning offered by the histogram mode, spectral lines were formed from time bins containing the signal from the fingertip only, so that the reflection artefacts from the glass slide do not appear in the images. The applied exposure time was 1 ms.

Appendix F

List of Publications

Publication with primary contribution

- Kufcsák, A., Erdogan, A., Walker, R., Ehrlich, K., Tanner, M., Megia-Fernandez, A., Scholefield, E., Emanuel, P., Dhaliwal, K., Bradley, M., Henderson, R. K., and Krstajič, N. Time-resolved spectroscopy at 19,000 lines per second using a CMOS SPAD line array enables advanced biophotonics applications. Optics Express, 25(10):11103, may 2017. ISSN 1094-4087. doi: 10.1364/OE.25.011103

Publications with secondary contribution

- Ehrlich, K., Kufcsák, A., Krstajič, N., Henderson, R. K., Thomson, R. R., and Tanner, M. G. Fibre optic time-resolved spectroscopy using CMOS-SPAD arrays. volume 10058, page 100580H. International Society for Optics and Photonics, feb 2017a. ISBN 9781510605572. doi: 10.1117/12.2252090.
- Ehrlich, K., Kufcsák, A., McAughtrie, S., Fleming, H., Krstajic, N., Campbell, C. J., Henderson, R. K., Dhaliwal, K., Thomson, R. R., and Tanner, M. G. pH sensing through a single optical fibre using SERS and CMOS SPAD line arrays. Optics Express, 25(25):30976, dec 2017b. ISSN 1094-4087. doi: 10.1364/OE.25.030976.
- Ehrlich, K., Kufcsák, A., McAughtrie, S., Fleming, H., Krstajič c, N., Campbell, C. J., Henderson, R. K., Dhaliwal, K., Thomson, R. R., and Tanner, M. G. Time-resolved single photon spectroscopy through a single optical fibre for miniaturised medical probe design. In Popp, J., Tuchin, V. V., and Pavone, F. S., editors, Biophotonics: Photonic Solutions for Better Health Care VI, volume 10685, page 24. SPIE, may 2018. ISBN 9781510618961. doi: 10.1117/12.2307334

Research Article | Vol. 25, No. 10 | 15 May 2017 | OPTICS EXPRESS 11103

Optics EXPRESS

Time-resolved spectroscopy at 19,000 lines per second using a CMOS SPAD line array enables advanced biophotonics applications

A. KUFCSÁK,[1,5] A. ERDOGAN,[1] R. WALKER,[1] K. EHRLICH,[2,3] M. TANNER,[2,3] A. MEGIA-FERNANDEZ,[3,4] E. SCHOLEFIELD,[3] P. EMANUEL,[3] K. DHALIWAL,[3] M. BRADLEY,[3,4] R. K. HENDERSON,[1] AND N. KRSTAJIĆ[1,3,*]

[1]*Institute for Integrated Micro and Nano Systems, School of Engineering, University of Edinburgh, King's Buildings, Alexander Crum Brown Road, Edinburgh EH9 3FF, UK*
[2]*Scottish Universities Physics Alliance (SUPA), Institute of Photonics and Quantum Science, Heriot-Watt University, Edinburgh EH14 4AS, UK*
[3]*EPSRC IRC Hub in Optical Molecular Sensing & Imaging, Centre for Inflammation Research, Queen's Medical Research Institute, University of Edinburgh, 47 Little France Crescent, Edinburgh EH16 4TJ, UK*
[4]*School of Chemistry, EaStChem, University of Edinburgh, Joseph Black Building, West Mains Road, Edinburgh EH9 3FJ, UK*
[5]*a.kufcsak@ed.ac.uk*
**n.krstajic@physics.org*

Abstract: A SPAD-based line sensor fabricated in 130 nm CMOS technology capable of acquiring time-resolved fluorescence spectra (TRFS) in 8.3 milliseconds is presented. To the best of our knowledge, this is the fastest time correlated single photon counting (TCSPC) TRFS acquisition reported to date. The line sensor is an upgrade to our prior work and incorporates: i) parallelized interface from sensor to surrounding circuitry enabling high line rate to the PC (19,000 lines/s) and ii) novel time-gating architecture where detected photons in the OFF region are rejected digitally after the output stage of the SPAD. The time-gating architecture was chosen to avoid electrical transients on the SPAD high voltage supplies when gating is achieved by excess bias modulation. The time-gate has an adjustable location and time window width allowing the user to focus on time-events of interest. On-chip integrated center-of-mass (CMM) calculations provide efficient acquisition of photon arrivals and direct lifetime estimation of fluorescence decays. Furthermore, any of the SPC, TCSPC and on-chip CMM modes can be used in conjunction with the time-gating. The higher readout rate and versatile architecture greatly empower the user and will allow widespread applications across many techniques and disciplines. Here we focused on 3 examples of TRFS and time-gated Raman spectroscopy: i) kinetics of chlorophyll A fluorescence from an intact leaf; ii) kinetics of a thrombin biosensor FRET probe from quenched to fluorescence states; iii) *ex vivo* mouse lung tissue autofluorescence TRFS; iv) time-gated Raman spectroscopy of toluene at 3056 cm^{-1} peak. To the best of our knowledge, we detect spectrally for the first time the fast rise in fluorescence lifetime of chlorophyll A in a measurement over *single* fluorescent transient.

OCIS codes: (040.1240) Arrays; (120.6200) Spectrometers and spectroscopic instrumentation; (300.2530) Fluorescence, laser-induced; (300.6450) Spectroscopy, Raman; (300.6500) Spectroscopy, time-resolved.

References and Links

1. L. M. Bollinger and G. E. Thomas, "Measurement of the time dependence of scintillation intensity by a delayed-coincidence method," Rev. Sci. Instrum. **32**(9), 1044–1050 (1961).
2. W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques* (Springer, 2005).
3. D. Tyndall, B. R. Rae, D. D.-U. Li, J. Arlt, A. Johnston, J. A. Richardson, and R. K. Henderson, "A high-throughput time-resolved mini-silicon photomultiplier with embedded fluorescence lifetime estimation in 0.13 µm CMOS," IEEE Trans. Biomed. Circuits Syst. **6**(6), 562–570 (2012).

#285077 https://doi.org/10.1364/OE.25.011103
Journal © 2017 Received 20 Jan 2017; revised 4 Apr 2017; accepted 9 Apr 2017; published 4 May 2017

Optics EXPRESS

4. C. Veerappan, J. Richardson, R. Walker, D.-U. Li, M. W. Fishburn, Y. Maruyama, D. Stoppa, F. Borghetti, M. Gersbach, R. K. Henderson, and E. Charbon, "A 160 x 128 single-photon image sensor with on-pixel 55 ps 10 bit time-to-digital converter," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International* (2011), pp. 312–314.
5. E. Charbon, "Single-photon imaging in complementary metal oxide semiconductor processes," Philos. Trans. Roy. Soc. London A **372**, 1-31 (2014).
6. R. J. Strasser, M. Tsimilli-Michael, and A. Srivastava, "Analysis of the Chlorophyll a Fluorescence Transient," in *Chlorophyll a Fluorescence*, G. C. Papageorgiou and Govindjee, eds. (Springer Netherlands, 2004), pp. 321–362.
7. H. M. Kalaji, G. Schansker, R. J. Ladle, V. Goltsev, K. Bosa, S. I. Allakhverdiev, M. Brestic, F. Bussotti, A. Calatayud, P. Dąbrowski, N. I. Elsheery, L. Ferroni, L. Guidi, S. W. Hogewoning, A. Jajoo, A. N. Misra, S. G. Nebauer, S. Pancaldi, C. Penella, D. Poli, M. Pollastrini, Z. B. Romanowska-Duda, B. Rutkowska, J. Serôdio, K. Suresh, W. Szulc, E. Tambussi, M. Yanniccari, and M. Zivcak, "Frequently asked questions about in vivo chlorophyll fluorescence: practical issues," Photosynth. Res. **122**(2), 121–158 (2014).
8. K. Suhling, L. M. Hirvonen, J. A. Levitt, P.-H. Chung, C. Tregidgo, A. Le Marois, D. A. Rusakov, K. Zheng, S. Ameer-Beg, S. Poland, S. Coelho, R. Henderson, and N. Krstajic, "Fluorescence lifetime imaging (FLIM): Basic concepts and some recent developments," Med. Photonics **27**, 3–40 (2015).
9. S. P. Poland, N. Krstajić, S. Coelho, D. Tyndall, R. J. Walker, V. Devauges, P. E. Morton, N. S. Nicholas, J. Richardson, D. D.-U. Li, K. Suhling, C. M. Wells, M. Parsons, R. K. Henderson, and S. M. Ameer-Beg, "Time-resolved multifocal multiphoton microscope for high speed FRET imaging in vivo," Opt. Lett. **39**(20), 6013–6016 (2014).
10. S. P. Poland, N. Krstajić, J. Monypenny, S. Coelho, D. Tyndall, R. J. Walker, V. Devauges, J. Richardson, N. Dutton, P. Barber, D. D.-U. Li, K. Suhling, T. Ng, R. K. Henderson, and S. M. Ameer-Beg, "A high speed multifocal multiphoton fluorescence lifetime imaging microscope for live-cell FRET imaging," Biomed. Opt. Express **6**(2), 277–296 (2015).
11. M. Popleteeva, K. T. Haas, D. Stoppa, L. Pancheri, L. Gasparini, C. F. Kaminski, L. D. Cassidy, A. R. Venkitaraman, and A. Esposito, "Fast and simple spectral FLIM for biochemical and medical imaging," Opt. Express **23**(18), 23511–23525 (2015).
12. N. Krstajić, J. Levitt, S. Poland, S. Ameer-Beg, and R. Henderson, "256 × 2 SPAD line sensor for time resolved fluorescence spectroscopy," Opt. Express **23**(5), 5653–5669 (2015).
13. K. Ehrlich, A. Kufcsák, N. Krstajić, R. K. Henderson, R. R. Thomson, and M. G. Tanner, "Fibre optic time-resolved spectroscopy using CMOS-SPAD arrays," Proc. SPIE **1005**, 100580H (2017).
14. J. M. Beechem, "Picosecond fluorescence decay curves collected on millisecond time scale: direct measurement of hydrodynamic radii, local/global mobility, and intramolecular distances during protein-folding reactions," Methods Enzymol. **278**, 24–49 (1997).
15. S. V. Kathuria and O. Bilsel, "Probing microsecond reactions with microfluidic mixers and TCSPC," in *Advanced Time-Correlated Single Photon Counting Applications*, W. Becker, ed., Springer Series in Chemical Physics No. 111 (Springer International Publishing, 2015), pp. 357–384.
16. W. Becker, V. Shcheslavkiy, S. Frere, and I. Slutsky, "Spatially resolved recording of transient fluorescence-lifetime effects by line-scanning TCSPC," Microsc. Res. Tech. **77**(3), 216–224 (2014).
17. D. R. Yankelevich, D. Elson, and L. Marcu, "Pulse sampling technique," in *Fluorescence Lifetime Spectroscopy and Imaging* (CRC, 2014), pp. 87–102.
18. H. Xie, J. Bec, J. Liu, Y. Sun, M. Lam, D. R. Yankelevich, and L. Marcu, "Multispectral scanning time-resolved fluorescence spectroscopy (TRFS) technique for intravascular diagnosis," Biomed. Opt. Express **3**(7), 1521–1533 (2012).
19. D. R. Yankelevich, D. Ma, J. Liu, Y. Sun, Y. Sun, J. Bec, D. S. Elson, and L. Marcu, "Design and evaluation of a device for fast multispectral time-resolved fluorescence spectroscopy and imaging," Rev. Sci. Instrum. **85**(3), 034303 (2014).
20. J. Ervin, J. Sabelko, and M. Gruebele, "Submicrosecond real-time fluorescence sampling: application to protein folding," J. Photochem. Photobiol. B **54**(1), 1–15 (2000).
21. E. B. Ishay, G. Hazan, G. Rahamim, D. Amir, and E. Haas, "An instrument for fast acquisition of fluorescence decay curves at picosecond resolution designed for "double kinetics" experiments: Application to fluorescence resonance excitation energy transfer study of protein folding," Rev. Sci. Instrum. **83**(8), 084301 (2012).
22. D. McLoskey, D. Campbell, A. Allison, and G. Hungerford, "Fast time-correlated single-photon counting fluorescence lifetime acquisition using a 100 MHz semiconductor excitation source," Meas. Sci. Technol. **22**(6), 67001 (2011).
23. L. Marcu, "Fluorescence lifetime techniques in medical applications," Ann. Biomed. Eng. **40**(2), 304–331 (2012).
24. Govindje, "Sixty-three years since Kautsky: chlorophyll a fluorescence," Funct. Plant Biol. **22**, 131–160 (1995).
25. K. Maxwell and G. N. Johnson, "Chlorophyll fluorescence-a practical guide," J. Exp. Bot. **51**(345), 659–668 (2000).
26. A. Le Marois, S. Labouesse, K. Suhling, and R. Heintzmann, "Noise-corrected principal component analysis of fluorescence lifetime imaging data," J. Biophoton. (2016).

Optics EXPRESS

27. M. I. Rowley, A. C. C. Coolen, B. Vojnovic, and P. R. Barber, "Robust Bayesian fluorescence lifetime estimation, decay model selection and instrument response determination for low-intensity FLIM imaging," PLoS One **11**(6), e0158404 (2016).
28. D. Renker, "Geiger-mode avalanche photodiodes, history, properties and problems," Nucl. Instrum. Methods Phys. Res. **567**(1), 48–56 (2006).
29. A. Rochas, M. Gosch, A. Serov, P. A. Besse, R. S. Popovic, T. Lasser, and R. Rigler, "First fully integrated 2-D array of single-photon detectors in standard CMOS technology," IEEE Photonics Technol. Lett. **15**(7), 963–965 (2003).
30. L. Pancheri and D. Stoppa, "A SPAD-based pixel linear array for high-speed time-gated fluorescence lifetime imaging," in *Proceedings of ESSCIRC, 2009. ESSCIRC '09* (2009), pp. 428–431.
31. S.-S. Kiwanuka, T. K. Laurila, J. H. Frank, A. Esposito, K. Blomberg von der Geest, L. Pancheri, D. Stoppa, and C. F. Kaminski, "Development of broadband cavity ring-down spectroscopy for biomedical diagnostics of liquid analytes," Anal. Chem. **84**(13), 5489–5493 (2012).
32. J. M. Pavia, M. Scandini, S. Lindner, M. Wolf, and E. Charbon, "A 1 x 400 backside-illuminated SPAD sensor with 49.7 ps resolution, 30 pJ/sample TDCs fabricated in 3D CMOS technology for near-infrared optical tomography," IEEE J. Solid-State Circuits **50**(10), 2406–2418 (2015).
33. S. Burri, H. Homulle, C. Bruschini, and E. Charbon, "LinoSPAD: a time-resolved 256×1 CMOS SPAD line sensor system featuring 64 FPGA-based TDC channels running at up to 8.5 giga-events per second," Proc. SPIE **9899**, 98990D (2016).
34. Y. Maruyama, J. Blacksberg, and E. Charbon, "A 1024 x 8, 700-ps time-gated SPAD line sensor for planetary surface exploration with laser Raman spectroscopy and LIBS," IEEE J. Solid-State Circuits **49**(1), 179–189 (2014).
35. J. Blacksberg, Y. Maruyama, E. Charbon, and G. R. Rossman, "Fast single-photon avalanche diode arrays for laser Raman spectroscopy," Opt. Lett. **36**(18), 3672–3674 (2011).
36. J. Kostamovaara, J. Tenhunen, M. Kögler, I. Nissinen, J. Nissinen, and P. Keränen, "Fluorescence suppression in Raman spectroscopy using a time-gated CMOS SPAD," Opt. Express **21**(25), 31632–31645 (2013).
37. Z. Li and M. J. Deen, "Towards a portable Raman spectrometer using a concave grating and a time-gated CMOS SPAD," Opt. Express **22**(15), 18736–18747 (2014).
38. I. Nissinen, J. Nissinen, P. Keränen, A. K. Länsman, J. Holma, and J. Kostamovaara, "A 2×(4)×128 Multitime-Gated SPAD Line Detector for Pulsed Raman Spectroscopy," IEEE Sens. J. **15**, 1358–1365 (2015).
39. I. Nissinen, J. Nissinen, P. Keränen, and J. Kostamovaara, "On the effects of the time gate position and width on the signal-to-noise ratio for detection of Raman spectrum in a time-gated CMOS single-photon avalanche diode based sensor," Sens. Actuators B Chem. (n.d.).
40. F. M. Rocca, J. Nedbal, D. Tyndall, N. Krstajić, D. D.-U. Li, S. M. Ameer-Beg, and R. K. Henderson, "Real-time fluorescence lifetime actuation for cell sorting using a CMOS SPAD silicon photomultiplier," Opt. Lett. **41**(4), 673–676 (2016).
41. S. P. Poland, A. T. Erdogan, N. Krstajić, J. Levitt, V. Devauges, R. J. Walker, D. D.-U. Li, S. M. Ameer-Beg, and R. K. Henderson, "New high-speed centre of mass method incorporating background subtraction for accurate determination of fluorescence lifetime," Opt. Express **24**(7), 6899–6915 (2016).
42. D.-U. Li, B. Rae, R. Andrews, J. Arlt, and R. Henderson, "Hardware implementation algorithm and error analysis of high-speed fluorescence lifetime sensing systems using center-of-mass method," J. Biomed. Opt. **15**, 17006 (2010).
43. D. D.-U. Li, J. Arlt, D. Tyndall, R. Walker, J. Richardson, D. Stoppa, E. Charbon, and R. K. Henderson, "Video-rate fluorescence lifetime imaging camera with CMOS single-photon avalanche diode arrays and high-speed imaging algorithm," J. Biomed. Opt. **16**(9), 096012 (2011).
44. A. Kufcsák and N. Krstajić, "Datasets for TRFS Optics Express paper Kufcsák et al, link http://dx.doi.org/10.7488/ds/2000," (2017).
45. D. E. Battey, J. B. Slater, R. Wludyka, H. Owen, D. M. Pallister, and M. D. Morris, "Axial transmissive f/1.8 imaging Raman spectrograph with volume-phase holographic filter and grating," Appl. Spectrosc. **47**(11), 1913–1919 (1993).
46. R. E. Bell, "Exploiting a transmission grating spectrometer," Rev. Sci. Instrum. **75**(10), 4158–4161 (2004).
47. N. Krstajić, C. T. A. Brown, K. Dholakia, and M. E. Giardini, "Tissue surface as the reference arm in Fourier domain optical coherence tomography," J. Biomed. Opt. **17**(7), 071305 (2012).
48. S. Preus, "DecayFit - fluorescence decay analysis software 1.3, FluorTools, www.fluortools.com," (2014).
49. S. Preus, K. Kilså, F.-A. Miannay, B. Albinsson, and L. M. Wilhelmsson, "FRETmatrix: a general methodology for the simulation and analysis of FRET in nucleic acids," Nucleic Acids Res. **41**(1), e18 (2013).
50. M. Bradley, S. V. Chankeshwara, and A. Megia-Fernandez, "Optical probe for thrombin," U.S. patent WO2016151297 A1 (September 29, 2016).
51. R. H. Clarke, S. Londhe, W. R. Premasiri, and M. E. Womble, "Low-resolution Raman spectroscopy: instrumentation and applications in chemical analysis," J. Raman Spectrosc. **30**(9), 827–832 (1999).
52. F. Franck, D. Dewez, and R. Popovic, "Changes in the room-temperature emission spectrum of chlorophyll during fast and slow phases of the Kautsky effect in intact leaves," Photochem. Photobiol. **81**(2), 431–436 (2005).
53. J. R. Lakowicz, H. Szmacinski, and M. L. Johnson, "Calcium imaging using fluorescence lifetimes and long-wavelength probes," J. Fluoresc. **2**(1), 47–62 (1992).

Optics EXPRESS

54. T. Terai and T. Nagano, "Small-molecule fluorophores and fluorescent probes for bioimaging," Pflugers Arch. **465**(3), 347–359 (2013).
55. B. Mills, M. Bradley, and K. Dhaliwal, "Optical imaging of bacterial infections," Clin. Transl. Imaging **4**(3), 163–174 (2016).
56. J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. K. Henderson, "A 32 x32 50ps resolution 10 bit time to digital converter array in 130nm CMOS for time correlated imaging," in *IEEE Custom Integrated Circuits Conference, 2009. Cicc '09* (2009), pp. 77–80.
57. A. Erdogan, R. Walker, N. Finlayson, N. Krstajic, G. Williams, and R. Henderson, "A 16.5 Giga Events/s 1024 × 8 SPAD Line Sensor with per-pixel Zoomable 50ps-6.4ns/bin Histogramming TDC," in *VLSI Symposium* (IEEE, 2017).
58. J. Arlt, D. Tyndall, B. R. Rae, D. D.-U. Li, J. A. Richardson, and R. K. Henderson, "A study of pile-up in integrated time-correlated single photon counting systems," Rev. Sci. Instrum. **84**(10), 103105 (2013).
59. O. Holub, M. J. Seufferheld, C. Gohlke, G. J. Heiss, and R. M. Clegg, "Fluorescence lifetime imaging microscopy of Chlamydomonas reinhardtii: non-photochemical quenching mutants and the effect of photosynthetic inhibitors on the slow chlorophyll fluorescence transient," J. Microsc. **226**(2), 90–120 (2007).
60. U. Noomnarm and R. M. Clegg, "Fluorescence lifetimes: fundamentals and interpretations," Photosynth. Res. **101**(2-3), 181–194 (2009).
61. T. A. Abbas, N. A. W. Dutton, O. Almer, S. Pellegrini, Y. Henrion, and R. K. Henderson, "Backside illuminated SPAD image sensor with 7.83μm pitch in 3D-Stacked CMOS technology," in *62nd International Electronic Devices Meeting* (IEEE, 2016).

1. Introduction

The unique advantage of single photon detection is that each photon can be treated as an individual, digital event. As the early pioneers of time-resolved detection Bollinger and Thomas [1] realised, single photon detection gives an experimenter immense flexibility. For example, as first shown in [1], it is possible to measure the time between the light pulse interacting with the sample and the detected luminescence (or scattered) photon arrival time. When each photon is time-correlated with the respective triggered laser and sorted into a histogram the technique is known as time-correlated single photon counting (TCSPC, see Table 1 for all abbreviations).

Table 1. Abbreviations used in the article

CMM	Center-of-mass method, often called method of moments
CMOS	Complementary metal-oxide-semiconductor
DCR	Dark count rate
FAM	5-carboxyfluorescein
FLIM	Fluorescence lifetime imaging
FPGA	Field programmable gate array
FRET	Förster resonance energy transfer
FWHM	Full width half maximum
IIR	Infinite impulse response (filter)
IRF	Instrument response function
LLTF	Laser line tunable filter
LRRS	Low resolution Raman Spectroscopy
MMP	Matrix metalloproteinase
NIM	Nuclear instrumentation module (electrical signaling standard)
PCB	Printed circuit board
SPAD	Single photon avalanche diodes
SPC	Single photon counting
TCSPC	Time correlated single photon counting
TDC	Time to digital converter
TRFS	Time-resolved fluorescence spectroscopy
TTL	Transistor-transistor logic (electrical signaling standard)

Typically, TCSPC is a slow technique and is limited only to low light scenarios. Imposition of the detector dead-time and photon statistics constraints (pile-up) usually do not affect the inherently photon-starved measurements (see Chapter 1 in [2]), but they do slow down techniques where photon budgets are not so constrained. The slowness of TCSPC is mostly due to the fact that detectors in most TCSPC environments are photo-multiplier tubes (PMTs) which are hard to array, resulting in most setups relying on just one or two detectors.

Optics EXPRESS

Our aim in this paper is to show that recent advances in complementary metal-oxide-semiconductor (CMOS) technology and single photon avalanche diodes (SPADs) allow massively-parallelized photon sorting [3–5] enabling both low light and light bursts to be efficiently time-stamped spectrally in a *single fluorescence transient* experiment. We use the term fluorescence transient to mean a change in fluorescence intensity [6,7] unrelated to excited state (fluorescence or phosphorescence) lifetime.

While most recent applications of CMOS SPAD arrays have focused on fast fluorescence lifetime imaging (FLIM) [8] using TCSPC [4,9–11], here we build on our prior work in time-resolved CMOS SPAD line sensors [12,13] to show that fluorescence transients can be measured both in the millisecond and nanoseconds regime ("double kinetics" [14,15]) in a single experimental sequence and in the spectral domain. Previously, to study fast fluorescent transients using TCSPC in the millisecond regime the transients needed to be induced many times in order to acquire time-resolved fluorescence kinetics [14,16]. Direct decay measurements have been achieved by a pulse sampling technique [17] and applied in tissue autofluorescence studies [18,19] and protein folding [14,20,21]. Pulse sampling achieves nanosecond resolution, but requires a high-end oscilloscope (2 GHz or more), while the pulsed lasers deployed tend to have high peak power to enable fast and direct detection of the fluorescence decay. Therefore, one of the significant advantages of the work presented here is that we demonstrate time-resolved spectroscopy using high repetition rate picosecond laser diodes [22] with low average power (<0.5 mW) making it immediately more attractive for clinical use [23] as well as opening up a wider range of other applications due to the lower cost and reduced number of safety issues.

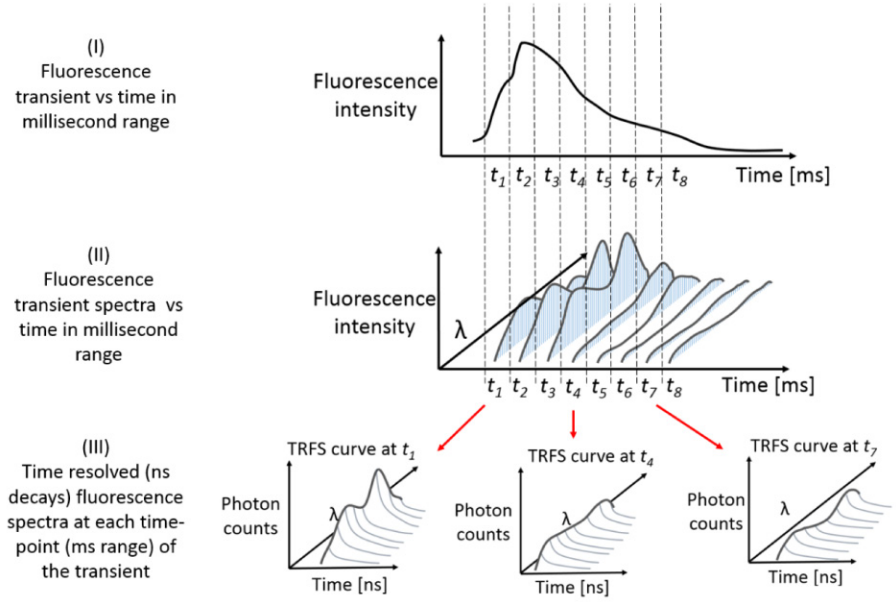

Fig. 1. Fluorescence transients can easily be measured using a photodiode in the microsecond to millisecond regime (top graph (I)). Recent developments in compact spectrometers have advanced this further allowing a spectral view into fluorescence transient evolution (middle graph (II)). In this paper we are able to show spectral "double kinetics" in nanosecond and millisecond regime whereby the transient is induced only once. For each time point (t_1, t_2, ...) a TRFS data cube is acquired during the single transient.

Figure 1 shows the outline of our proposed luminescence transient analysis. The top graph (I) shows a fluorescence transient ("single kinetics") from an analogue photo-detector such as a photodiode. This is easily achievable and the next step is either to have a fluorescence decay

Optics EXPRESS

for each time point [14–16] or spectrum for each time point as shown in middle graph (II). Our work advances the spectral lifetime acquisition or time-resolved fluorescence spectroscopy (TRFS) and we demonstrate "double kinetics" in the spectral domain potentially exploitable in a large number of phenomena such as chlorophyll A fluorescence (plant fluorescence), fluorescence quenching, photo-bleaching, Förster resonance energy transfer (FRET) and protein folding dynamics to name but a few. We particularly pay attention to plant fluorescence transients as an important example of fluorescence kinetics. As described in numerous introductory reviews [7,24,25], most of the plant fluorescence is attributed to chlorophyll A. Plant fluorescence is also an important indicator of plant stress and photosynthesis efficiency. Under controlled illumination fluorescence increases rapidly (during the first second, often 100 ms) which is followed by a slow fall (several minutes depending on plant species). This transient is usually referred to as the Kautsky effect. While the fluorescence properties of the slow fall have been studied in detail we demonstrate our sensor can be applied to studying the rapid rise in terms of spectral lifetime.

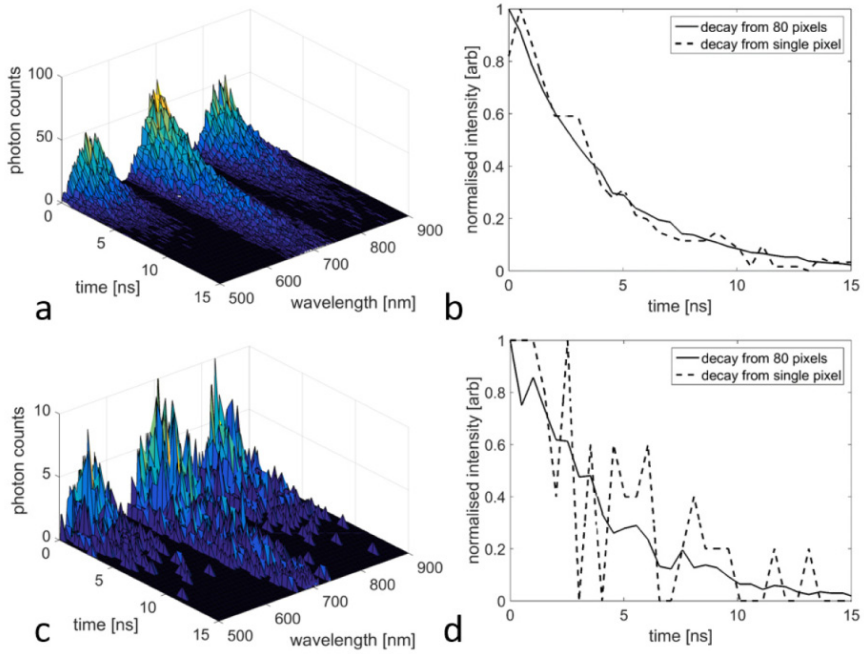

Fig. 2. Demonstration of spectral decays in two simulated scenarios. The first scenario is photon rich: (a) defined here as most pixels having more than 2000 counts in decay. Plot (b) shows the difference between the decay in single pixel and binning 80 pixels which results in a better looking decay. The second scenario is photon starved: (c) where the decay derived from a single pixel in (d) is too noisy. However, the decay from 80 binned pixels is good enough to allow the user to manipulate the data cube in an optimal way given the amount of light detected. Spectral decays were simulated in Matlab 2014a (Mathworks, USA), shot noise limited, visible to near-infrared TRFS data cubes. Spectral decays have 0.4 ns time resolution and 1.6 nm spectral resolution. Representing broadly the properties of the CMOS SPAD line sensor presented here.

Lastly, by binning the TRFS data cube either in the time or wavelength domain it can be adapted for time-resolved imaging. To illustrate this and similar scenarios, Fig. 2 shows two simulated TRFS data cubes in Figs. 2(a) and 2(c) and sample decays in Figs. 2(b) and 2(d) taken from the 650 nm - 750 nm region. Three emission peaks are simulated to have different decays to demonstrate a scenario where three fluorophore emissions are excited at 488 nm, 625 nm and 780 nm and studied simultaneously in a broadband visible to near-infrared

Optics EXPRESS

spectrometer. When TRFS at high spectral resolution is needed then binning can be done along the time axis. Similarly, if the time resolution needs to be kept, more counts can be gathered by binning in the spectral domain. In fact, spectral binning can be applied until one gets a single TCSPC decay curve. The compromise is that binning along time or wavelength axis entails loss of time or spectral resolution respectively, even though advanced strategies have been deployed recently [26,27]. The key aim is always to gather more counts in a decay, which in turn could enable better exponential fitting. We apply this methodology throughout our study of TRFS below.

Solid-state single photon detectors have a long history dating back to the 1960's [28]. Since the first integrated SPAD array was created using industrial CMOS technology [29], a variety of CMOS fabricated SPAD sensors have been designed. CMOS technology allows enhanced detection with additional logic on the detector device and enables cost-effective mass production of such sensors. Recent systems apply SPAD-based line sensors, where light from a point of the specimen is dispersed over several pixels. Timing information is either collected by time-gated photon counting [11,30,31], or by applying time-to-digital converters (TDCs) for each pixel, realizing a multi-channel TCSPC system [12]. Spectrally resolved lifetime images can be acquired by combining time-resolved line sensors with laser scanning microscopy [11].

SPAD line sensors have been reported targeting other applications, such as near-infrared optical tomography [32], 3D imaging based on time-of-flight measurements [33], laser-induced breakdown spectroscopy [34] and time-resolved Raman spectroscopy [35–39]. Line arrangements of pixels are in general favorable for increasing fill-factor of the sensor, as any additional logic can be placed next to the photon-sensitive area of the pixels. A line sensor implemented with 3D multi-wafer stacking CMOS technology has been demonstrated recently [32]. A novel approach in integrating flexible TDC architectures onto SPADs is to connect field programmable gate array (FPGA) based TDCs to SPADs, instead of integrating more functionality into the CMOS sensor (8.5 giga-events/s achieved) [33]. This approach separates analog SPAD circuits from digital FPGA side and allows users to benefit from recent FPGA advances.

TCSPC is restricted to process only one photon arrival per pre-defined exposure time [2]. The sensor we present enhances this by applying a separate TDC channel for each of its 256 pixels. To further increase the efficiency of time-stamping, time-resolved detection can also be performed in a center-of-mass (CMM) mode. This allows photon efficient processing of photon arrival times, because all processing is done on the sensor itself. This technique provides fast fluorescence decay analysis [40–43], while each operating mode of the sensor can be amended with an adjustable ON (allowing recording of photon events only during the time-gate) or OFF time-gate (register all photon events but the ones during the time-gate).

To demonstrate fast time-resolved spectroscopy using our 256 × 2 line sensor we studied: i) fluorescence kinetics of chlorophyll A in a Kautsky transient from an intact leaf [7,24,25] of *Viburnum rhytidophyllum*; ii) fluorescence kinetics of a linear FRET probe; iii) mouse lung tissue slide autofluorescence; iv) time-gated Raman spectroscopy of toluene. Our ambition was to demonstrate utility in a broad range of applications to stress-test our engineering implementation, but also to demonstrate how time-resolved CMOS SPAD line arrays lend themselves naturally to numerous applications in biomedical optics.

Optics EXPRESS

2. Methods

2.1 CMOS SPAD line sensor architecture

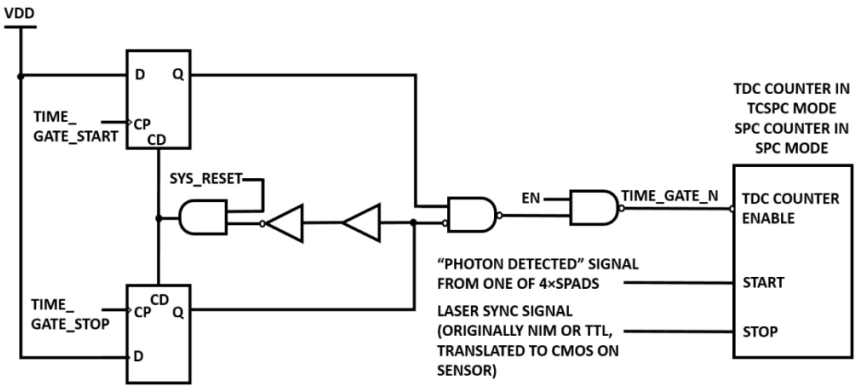

Fig. 3. Time-gated region is defined for each pixel by two global, pre-defined signals TIME_GATE_START and TIME_GATE_STOP. They define the time during which the TDC counter is enabled. TDC counter acts as TCSPC TDC counter in TCSPC mode, but in SPC mode it acts as a photon counter. Therefore the same circuit controls the time-gating behavior in both modes.

The common aspects with our prior work are as follows [12], the CMOS SPAD line array sensor comprises 2 rows of 256 pixels aiming for efficient photon detection at different wavelengths. Each pixel incorporates 4 SPADs optimized for blue spectral range and 4 SPADs optimized for red spectral range (referred to henceforth as blue SPADs and red SPADs) of 23.78 µm pitch yielding a 43.7% fill-factor. 256 × 26 bit TDC channels integrated on the sensor underpins TCSPC functionality. Detailed analysis of the instrument response function (IRF) and dark count rate results for blue and red SPADs are similar to our prior work [12] and are given in supplementary notes [44]. Three significant improvements were introduced in the work presented here: i) access to data lines coming from pixels was parallelized to increase the readout speed to 19,000 lines per second as opposed to 200 lines per second previously; ii) the time-gating circuit was optimized to work with the SPAD turned ON all the time, i.e. even outside the time-gate window; iii) a new PCB was designed to cater for the improved CMOS SPAD sensor, the firmware was rewritten to take advantage of both improved time-gating and parallelized access and finally, the front-end software was also completely re-written in Python (version 2.7.2) with both a graphical user interface and scripting options. Scripting allows the customization of the data acquisition protocol for specific requirements of any given experiment. For example, plant fluorescence displays particular kinetics where the rise in fluorescence intensity is fast (ms), while the fall is slow. Customizing the script allows combining TCSPC and CMM acquisitions and tailoring the parameters to obtain maximum information from the specimen. Compared to other TDC based CMOS SPAD line sensors, it has the highest line rate to the PC.

Optics EXPRESS

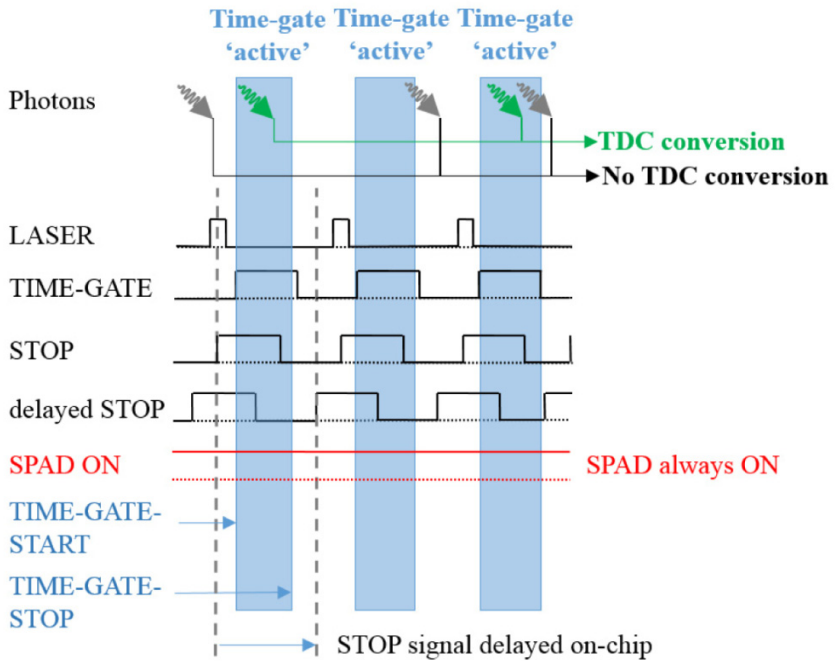

Fig. 4. Time-gated region (blue) is defined by TIME-GATE-START and TIME-GATE-STOP settings with respect to delayed STOP signal. As opposed to previous designs [12] the SPAD is always on. The TDC conversion only happens if photon arrival time takes place within the time-gated region. Photon arrivals outside the time-gated region do not result in TDC conversion, but they are electrically detected by the SPAD and hence there is a dead-time effect for subsequent detections.

The time-gating circuit is illustrated on Fig. 3. TIME_GATE_START and TIME_GATE_STOP are fed into clock pulse (CP) inputs of two D-flip-flops thus defining the time region during which TDC counter is enabled. As TDC counter acts as a reverse start-stop timing clock in TCSPC mode, it is started by photon detection and stopped by laser synchronization STOP signal (electrical TTL or NIM prior to reaching the PCB). The feedback into clear direct input (CD) reduces the risk of glitches. In SPC mode, TDC counter acts as a photon counter so the same time-gating circuit defines the regions during which the photon counting is performed. The timing diagram of the time-gating circuit is shown on Fig. 4 where LASER indicates a laser optical pulse, TIME-GATE is the time-gated region defined by TIME-GATE-START and TIME-GATE-STOP. STOP is the electrical timing pulse (TTL or NIM) received from the optical setup (for pulsed laser diodes STOP comes from the laser usually). STOP can be delayed on-chip and the time-gated region is defined with respect to this delayed version of STOP signal.

The time-gating can be applied in all modes: i) on-chip photon-efficient CMM; ii) SPC mode; or iii) TCSPC mode. In CMM, the time-gated region defines FIRST and LAST time bins required for accurate CMM calculation [42,43]. In time-gated photon counting the designated bin can potentially detect more photon events than in TCSPC mode where each exposure time allows only for one time-stamped photon event to be detected. Higher readout speed of 19,000 lines per second is the key enabler of new applications such as high speed TRFS presented below.

Optics EXPRESS

2.2 Sensor characterization – delay line and time-gating

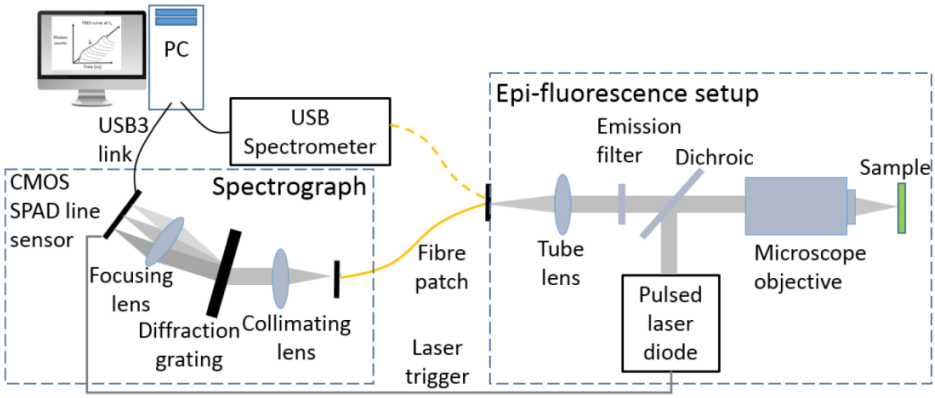

Fig. 5. Setup comprising epi-fluorescence light collection (right) and the spectrograph (left). The CMOS SPAD line sensor is placed in the focus of the spectrograph optics. Volume phase holographic grating is used to optimize light throughput. See Methods section for details.

The time-gating properties of the sensor were tested in ambient random light. The light was attenuated to avoid saturation of the pixels and early photon pile-up on the time histograms (photon arrival rate was less than 1% of the laser repetition rate). The 1 MHz laser sync output of the PLP10 laser driver was used as STOP for TCSPC. The position and width of the time-gate is based on a 128 step, on-chip delay line that is separate from the TDC circuits. The sensor allowed programmable delays for the rising and falling edge of the time-gate window with respect to the STOP signal using this delay line. The shortest applicable time-gate window was examined for both ON and OFF operations. Histograms of the gated photon arrival times in the ambient light were created with different gate positions. In both cases (time-gate ON and OFF) the full-width-half-maximum (FWHM) of the peaks with short delay time-gate (10 step delay for rising edge, 11 step delay for falling edge) were measured.

Time-gating was also tested in time-gated SPC mode. A 5 MHz triggering signal generated by the DG645 digital delay generator was connected both to the PLP10 laser driver and the CMOS SPAD line sensor using two separate channels. The first provided triggering of the laser pulses while the latter was used as STOP sync signal for gating. The delay between the two channels was controlled by the delay generator with 5 ps resolution. Scanning through a range of delays between the channels allowed us to move the specified time-gating window across the laser pulse.

2.3 TRFS from leaf, lung tissue, linear FRET probes and time-gated Raman spectroscopy

Figure 5 shows the CMOS SPAD line sensor characterization setup consisting of i) epifluorescence setup which collects fluorescence from the specimens and focuses it into multimode fiber (50 µm for leaf and FRET biosensor and 105 µm for lung slide, 0.22 NA multimode fiber, M14L01, M15L01, Thorlabs Ltd, UK) and ii) spectrograph consisting of dispersive optics with CMOS SPAD line sensor in focus. The specimens were placed in the focus of × 10 microscope (RMS10X, Thorlabs Ltd, UK) objective. Multiple pulsed laser sources were used. Leaf was illuminated by a 402 nm pulsed laser at 20 MHz repetition rate (PLP10, Hamamatsu, Japan). Peak power was 650 mW, pulse length was 54 ps resulting in an optimal 0.7 mW average output (assuming 20 MHz repetition rate). The illumination on the specimen was 0.4 mW due to the efficiency of the epi-fluorescence illumination path. The linear FRET probe were illuminated by a 483 nm pulsed laser diode at 20 MHz repetition rate. Peak power of 483 nm pulsed laser was 349 mW, pulse length was 60 ps resulting in a 0.42 mW average output. The illumination on the specimen was 0.27 mW. For alignment and

Optics EXPRESS

spectrograph resolutions tests a WhiteLaseMicro supercontinuum (NKT Photonics-Fianium, UK) was used. Supercontinuum was also used to illuminate the lung slide. Light above 700 nm was filtered out prior to illuminating the lung slide. The illumination on the specimen was 1.9 mW. Dichroic filters (FF414-DI01-25X36, Semrock, USA, for leaf and custom three color design from Chroma, USA for lung slide) were used to separate illumination from fluorescence. The spectrograph diffraction grating was volume phase holographic grating (600 lp/mm at 600nm, Wasatch Photonics) with collimating and focusing lens optimized for efficiency and spectral resolution [45–47]. Efficiency and wavelength span of the spectrograph were measured using a WhiteLaseMicro supercontinuum and a laser line tunable filter (LLTF Contrast, NKT Photonics,UK). For spectrograph efficiency measurement, the LLTF filter was set to 520 nm with a linewidth of 2.5 nm (verified using an Ocean Optics usb2000 spectrometer) and the optical power measured after the fiber input and after the focusing lens. The efficiency was measured to be 68%. Wavelength span and spectral resolution were verified by taking a spectrum of the fluorescent lamp using a custom spectrograph and an off-the-shelf spectrometer usb2000 (Ocean Optics, USA). To cover the 130 nm wavelength span across the 256 pixels of the line sensor the collimating lens used was a 50 mm focal length achromat (AC254-050-A-ML, Thorlabs, UK), while the focusing lens was 75 mm focal length achromat (AC508-075-A-ML). Optical spectral resolution was measured using fluorescence lamp and measured to be at least 2.5 nm. Fluorescence decays were fitted using non-linear least square fit routines extracted from DecayFit software [48,49].

The specimen for the leaf study was an intact leaf of *Viburnum rhytidophyllum* picked in December 2016 (location Edinburgh, UK) and illuminated for TRFS on the same day. The leaf was dark adapted for 1 hour in a closed box at room temperature upon removal from the tree. The fluorescence kinetics of the leaf was measured both in TCSPC and CMM modes. In TCSPC mode the applied exposure time was 8.3 µs which was followed by 42.9 µs gap due to firmware limitations. TRFS cubes were built from 1000 lines with TCSPC time-events with 8.3 ms overall exposure time over 51.2 ms with an additional gap of approximately 3 ms due to memory management on host computer. After the first 500 TRFS cubes data was collected with additional gaps of 100 ms between TRFS cubes to adapt for the kinetics of leaf fluorescence transient. In CMM mode the applied exposure time was 83.3 µs with an additional 0.25 µs gap due to firmware. Gap introduced by host computer was ~3 ms after every 1000 lines of CMM values. Accurate time point of each line was embedded into a header by the firmware.

The lung specimen slide originated from murine lungs (euthanized) embedded in paraffin wax following a standard protocol and sectioned at 5µm thickness on a microtome. The slides were then re-hydrated by immersing in xylene, 100%, 95%, 80%, 70% EtOH and dH_2O in that order. 10 µL of red Inspeck beads 1% (I-7224, Thermofisher Scientific, USA) were also inserted to allow for increased red fluorescence. They were then mounted under a coverslip in water soluble mountant. The exposure time in TCSPC mode was 8.3 µs.

A custom designed FRET molecular biosensor for thrombin [50] was used to study fluorescence kinetics. The compound consists of a linear peptide as a thrombin substrate with 5-carboxyfluorescein as a fluorophore on one end and methyl red as a quencher on the other, see Fig. 6. In the absence of thrombin, methyl red quenches 5-carboxyfluorescein fluorescence, but in presence of thrombin the enzymatic cleavage of the peptide releases the 5-carboxyfluorescein and the fluorescent signal increases. A solution containing the probe (5 µM) in matrix metalloproteinase (MMP) buffer (10mM CaCl2, 6.1 g Tris-HCl, 8.6 g NaCl per litre, pH 7.5) was incubated with thrombin (T9326-150UN, Sigma Aldrich, UK) (at a final concentration of 5 U/mL) in 5 × 2 mm microcuvette with a final volume of 300 µL (or in disposable plastic cuvettes with a final volume of 2 mL). Enzyme free reactions were used as a control. Where appropriate thrombin inhibitor Anti-thrombin III (AT3, Sigma Aldrich, UK) was pre-incubated at a final concentration of 0.4 µM. The enzyme was pipetted into cuvette 8

Optics EXPRESS

s after TRFS acquisition was started. The total TRFS scan was performed over 3 minutes 30 seconds. Firstly, 1000 TRFS cubes each with 8.3 ms exposure time over 54 ms elapsed time were acquired followed by 500 TRFS cubes each with 8.3 ms exposure time over 54 ms with an additional 100 ms gap between TRFS acquisition. Data acquisition parameters were selected to cater for fast transients in the beginning and slowed down for slower transients. This allowed us to study kinetics at short and long time scales. The initial delay of 8 s was used to acquire background fluorescence prior to enzyme activation. Diffusion effects were observed inside the cuvette.

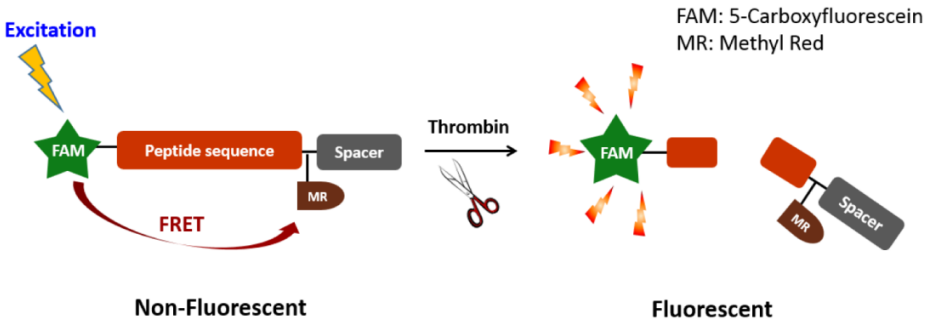

Fig. 6. FRET biosensor works by thrombin cleaving the peptide sequence separating the methyl red and 5-carboxyfluorescein (FAM) thus allowing FAM to fluoresce.

The TRFS data cubes for lung slide and custom FRET probe were post-processed by digital bandstop infinite impulse response (IIR) filter to remove the etaloning fringe observed during the experiment. As bare leaf experiments did not display etaloning, we believe this was induced by slide and cuvette. The bandstop frequency was observed in the power spectrum of the spectral curves and the IIR filter parameters were adjusted accordingly (filtering was implemented in Matlab 2014a). Decay data was used from original non-filtered data.

Time-gated Raman spectroscopy was achieved using the same configuration as for lung slide study above. A 10 mm cuvette filled with toluene was placed in front of the microscope objective. As the pulsed laser source had broad linewidth (7 nm), our implementation is in the low resolution Raman spectroscopy (LRSS) regime [51]. We use toluene as a strong Raman scatterer to test the time-gating operation of the circuit. We focus on 3056 cm^{-1} band (~567 nm for 483 nm excitation) to test the time-gating operation. An experiment was performed to acquire Raman spectra with and without time-gating (5.6 ns ON time-gate coinciding with the laser pulse) exposure time was set to 5 s and the laser repetiton rate was set to 50 MHz. Toluene sample was illuminated with 0.68 mW.

Due to firmware limitations, our minimum TCSPC exposure time was 8.3 µs. For each experiment, we maximized the number of TCSPC events detected by optimizing respective setups while taking into account both pile-up conditions (1%-10% of repetition rate limit for the number of TCSPC events detected) and counting loss distortions in TCSPC (see page 332 in [2]). To reduce the counting loss distortion the total number of TCSPC events detected on each pixel was set to be less than 50% of the maximum possible defined by the TCSPC exposure time.

Optics EXPRESS

3 Results

3.1 CMOS SPAD line sensor characterization - delay line and time-gating

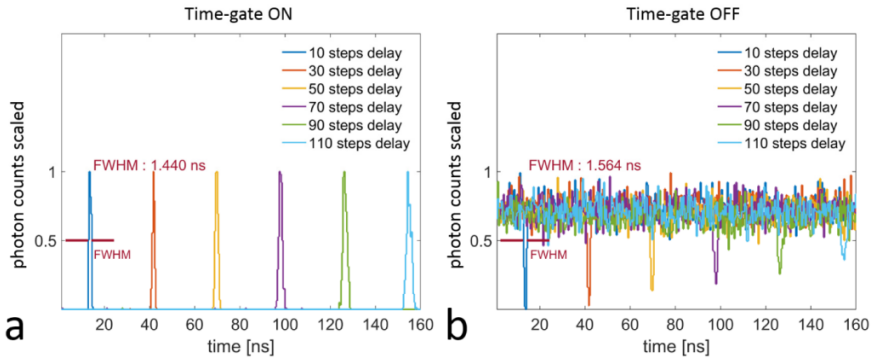

Fig. 7. Time-gated TCSPC histograms of photon arrival times in ambient light. Time-gates were positioned with an on-chip, 128 step delay line with respect to the STOP signal. The width of each time-gate was 1 step of the delay line. The average FWHM for a short delay (10 steps) ON time-gate was 1.44 ns (a). For the OFF time-gate it was 1.56 ns (b).

Figure 7(a) shows the shortest ON time-gate and Fig. 7(b) shows the shortest OFF time-gate for a representative pixel (pixel 100), both at different time positions on the histogram. The rising and falling edges of the time-gate window were adjusted using a 128 step on-chip delay line, with the zero delay on the rising edge causing this edge to be closest to the STOP signal. Longer delays in the window position resulted in wider time-gated regions, as a consequence of accumulated jitter of the steps in the delay line. The same accumulated jitter was responsible for having photons acquired even in the OFF region on the 'time-gate OFF' histograms. This can be avoided either by choosing a wider time-gate window, or by applying an external delay. The external delay can shift the histogram to a position where a certain time interval can be masked with a time-gate window that is closer to the STOP (i.e. lower on-chip delay). The average FWHM of the ON peak for 10 step delay is 1.44 ns. The average FWHM of the OFF peak for 10 step delay was 1.56 ns.

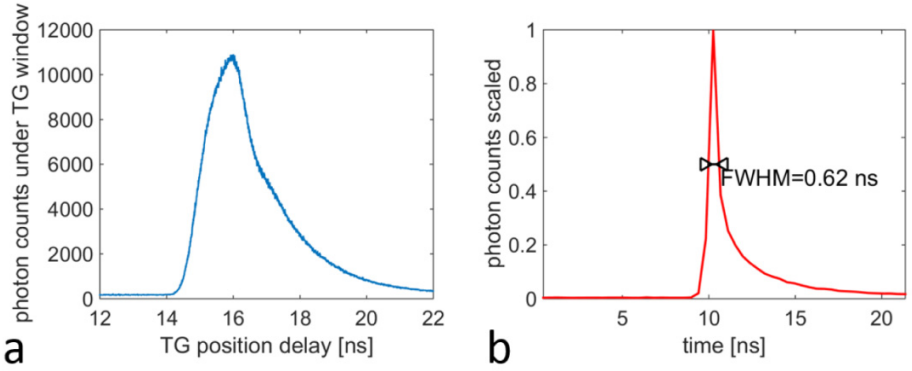

Fig. 8. Sweeping an external delay generator over a fixed time-gate window covering a laser pulse generates a more detailed picture of the time-gate (a) convolved with the asymmetric shapes of the laser pulse and red SPAD IRF (b). Blue SPADs were time-gated in (a) as its IRF does not have the diffuse tail present in red SPADs (b) (see supplementary notes [44]).

The resolution of the delay line was measured from the TDC resolution. The distance was calculated between the peaks of the histograms of 10 step and the 110 step delays. The number of TDC bins between the peaks multiplied with their resolution and divided by the

Optics EXPRESS

number of steps gives an estimated delay resolution, which was measured to be 1.4 ns. A finer delay can be applied in combination with an external delay generator.

Figure 8 shows the number of photons acquired in time-gated SPC mode with the minimum time-gate as a function of the applied delay with 5 ps resolution between the trigger and sync channels. The asymmetric shape of the resulting plot is due to the asymmetric shape of the detected laser pulses, and the fact that the described measurement results in a convolution of the laser peak with the time-gate window.

3.2 TRFS of leaf

Chlorophyll A transient is shown in Fig. 9 with a fast rise reached in less than 200 ms, while the slow fall in fluorescence intensity takes longer (tens of seconds). Figure 10(a) shows TRFS in the fast rise at 0.03 s, and at several time points in the slow fall, namely, 2.04 s in Fig. 10(b) and 120.04 s in Fig. 10(c). The TRFS curves at the beginning show a slight increase in lifetime during the fast rise while the snapshot at 120.04 s clearly demonstrates a decay in peak values of decays, but also a reduction in lifetime due to quenching. This is further demonstrated in Fig. 11(a), where an increase in fluorescence lifetime is visible during the initial rise. A clearer decrease in lifetime was detected during the slow fall in Fig. 11(b). The TRFS during the fast rise is available in Visualization 1 video, while the slow fall is in Visualization 2 video.

The time-resolved spectra shown in Fig. 10 have two peaks corresponding to 680 nm peak of Photosystem II and 740 nm peak of Photosystem I. The data obtained broadly correspond to prior data obtained using steady state spectrometers [52]. We also verified the spectral shape of leaf fluorescence with off-the-shelf steady-state usb2000 spectrometer (not shown).

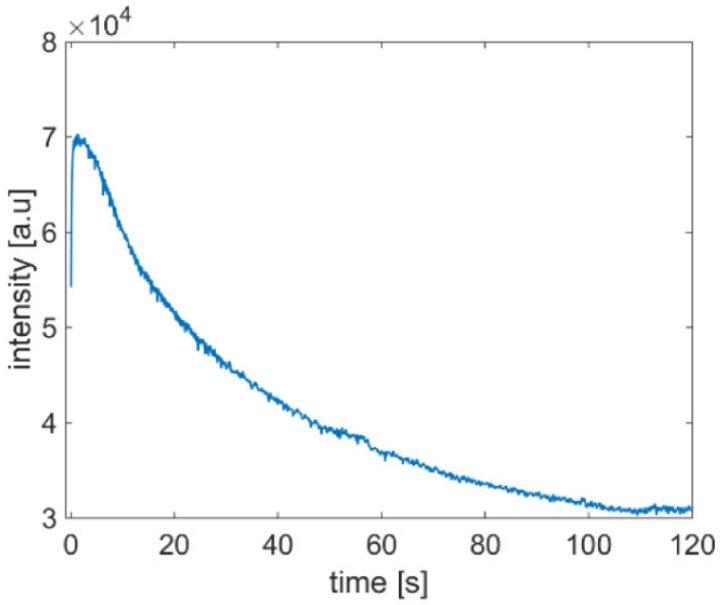

Fig. 9. Chlorophyll A transient curve from SPC acquisition over 120 s.

Optics EXPRESS

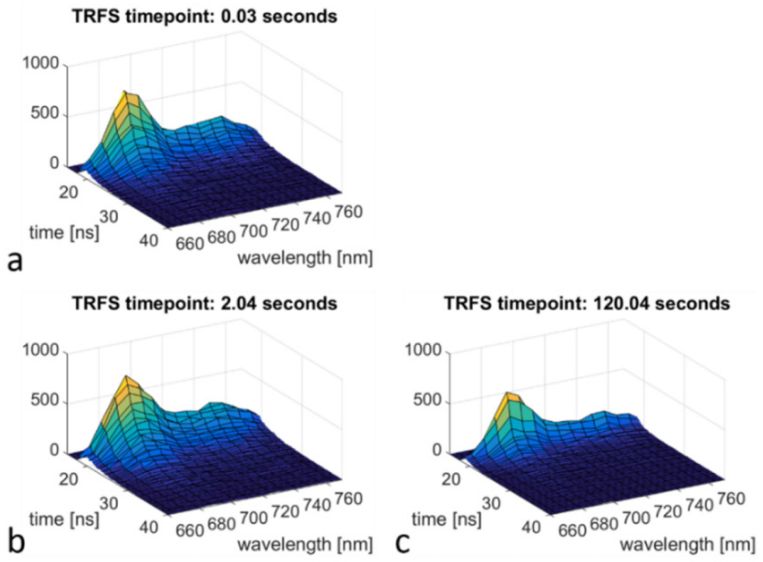

Fig. 10. Fluorescent transient of the leaf during fast rise and slow fall. Three TRFS time points at: (a). 0.03 s; (b). 2.04 s; and 120.04 s (c). See Visualization 1 and Visualization 2 for fast rise and slow fall videos respectively. 10 pixels were binned for each 3D plot in (a-b) resulting in 5 nm spectral coverage.

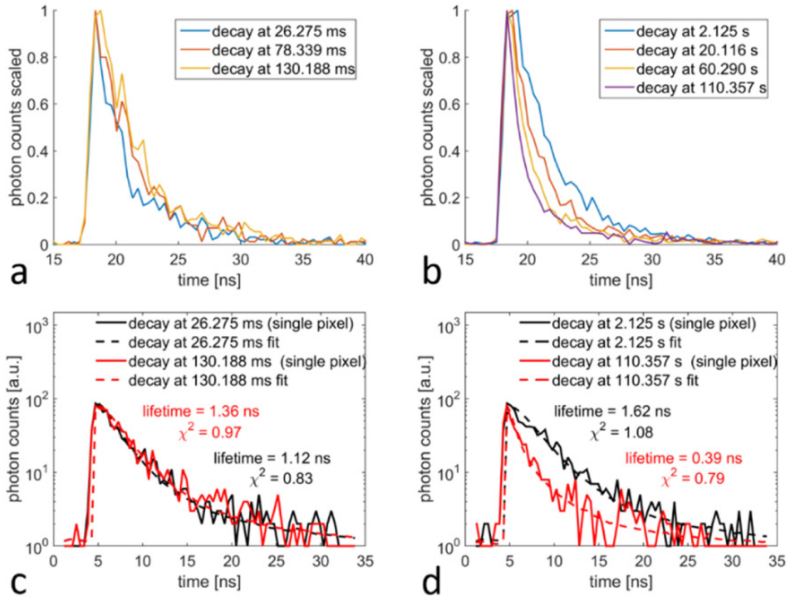

Fig. 11. Fluorescence decays extracted from TRFS data cubes taken during the fast rise (a,c) and the slow fall (b,d). Increase in lifetime on (a) is more obvious with decays fitted for 26.275 ms and 130.188 ms time points (c). Fluorescence quenching is indicative by reduction of lifetime in (b,d) as expected for the chlorophyll A transient. The decays in (a,b) were taken from single pixel (0.5 nm spectral coverage) and from 1 TRFS data cube (8.3 ms exposure time over 51.25 ms) at each time point during fast rise (a), 3 TRFS data cubes (25 ms exposure time over 153.75 ms) around each time point during slow fall (b).

Optics EXPRESS

In order to detect the lifetime increase during the fast rise, decays were binned over 150 lines acquired and over all pixels. This provides less noisy decays (Fig. 12) while maintaining reasonably good sampling in time. Spectral information is, however, lost. The lifetime increase during the fast fluorescence rise was also observed previously [16], but that measurement was done by periodic stimulation of chlorophyll A fluorescence kinetics, whereas we measure it in a single chlorophyll A fluorescent transient. Spectral lifetime information with fine sampling in time requires more efficient photon collection for which CMM mode of the chip was applied. CMM can provide lifetime estimates spectrally with much shorter exposure times and hence finer sampling in time. Figure 13 shows the lifetime rising over the first 40 ms of fast rise on the 680 nm and 740 nm peaks, using CMM mode. Several lines of CMM values from the sensor can be processed to have more counts over longer time and hence coarser sampling in time, or binning can be done over pixels reducing spectral resolution.

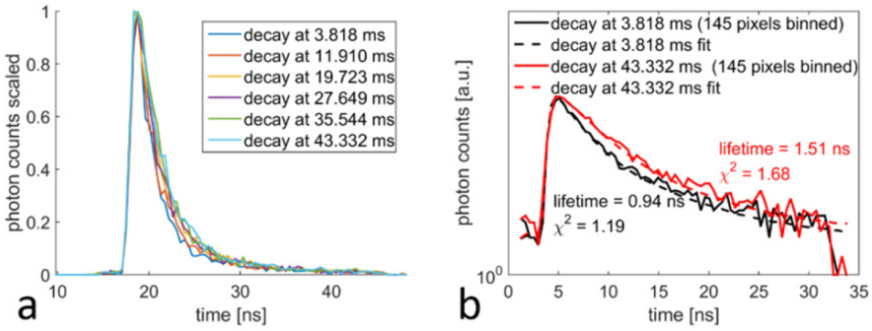

Fig. 12. Fluorescence decays during the fast rise (a). Lifetime change is more obvious when binning timestamps of all pixels, but no spectral information is available from decays in (a). The decays were created by binning 150 lines of TCSPC time-events of 1.3 ms total exposure time over 7.7 ms around each time point. The decay fits are shown in (b) for two time points.

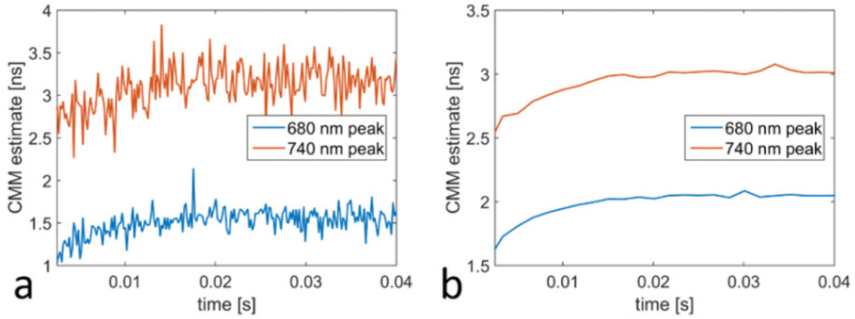

Fig. 13. Increasing fluorescence lifetime over the first 40 ms of fast rise on the 680 nm and 740 m peaks. CMM estimates of single pixel with photons captured over 166.6 µs shown on (a) broadly match lifetimes calculated by fitting decays to TCSPC data. Photons captured over 1.67 ms and 30 pixels result in smoother transients.

3.3 TRFS of thrombin FRET biosensor

Continuous TRFS acquisition before, during and after enzyme activation is illustrated in the figures below. Fluorescence intensity variation with time (extracted from TRFS) is shown in Fig. 14. The laser is switched on in the first second and the dip at 8 s indicates the time when the enzyme thrombin was pipetted into cuvette. The short dip in fluorescence intensity is likely to be due to movement of fluid during pipetting.

Optics EXPRESS

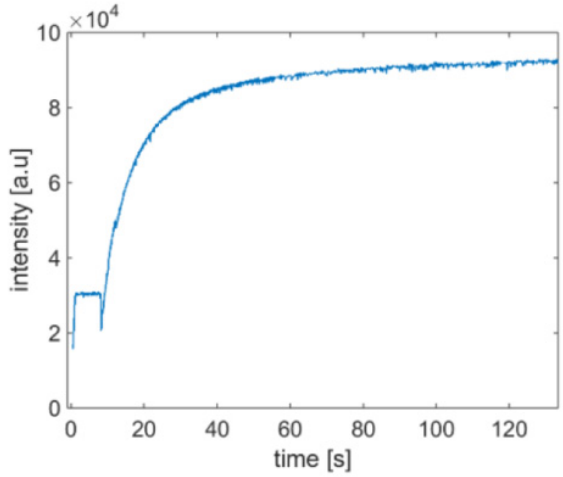

Fig. 14. Enzyme kinetics curve of a thrombin FRET biosensor. At 8 s thrombin was pipetted into a cuvette initiating a rise in fluorescence intensity. TRFS was acquired in 8.3 ms every 50 ms during the first 50 s (every 150 ms after 50s) and the kinetics curve was derived from underlying spectral double kinetics data.

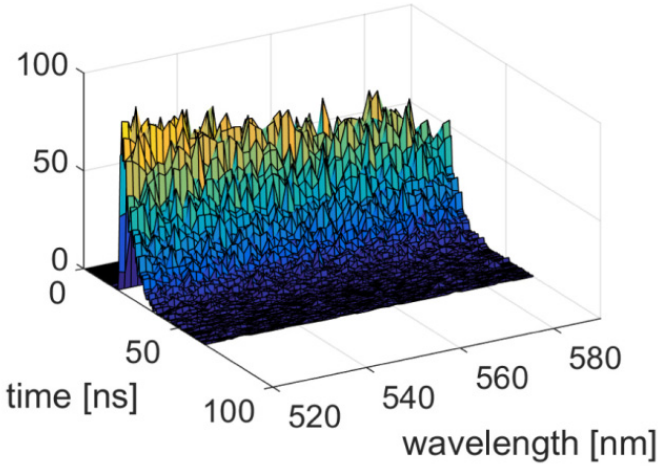

Fig. 15. Sample TRFS from the time point at 130 s. Spectral resolution is 0.4 nm and no binning was applied in the spectral domain.

The choice of acquisition parameters was led by memory requirements of 130 s acquisition. Sample TRFS data cube at full spectral resolution is shown on Fig. 15. Figure 16 shows decays extracted from the spectral peak (20 pixels, spectrum covered 8 nm) before and after enzyme activation. The lifetime change is minor indicating that background fluorescence is mostly due to unquenched 5-carboxyfluorescein. Increased time resolution and improved jitter will allow better decay analysis. The increase in fluorescence is due to 5-carboxyfluorescein being cleaved from the FRET biosensor. A similar effect has been observed for calcium dye Fluo-4am [16,53].

Optics EXPRESS

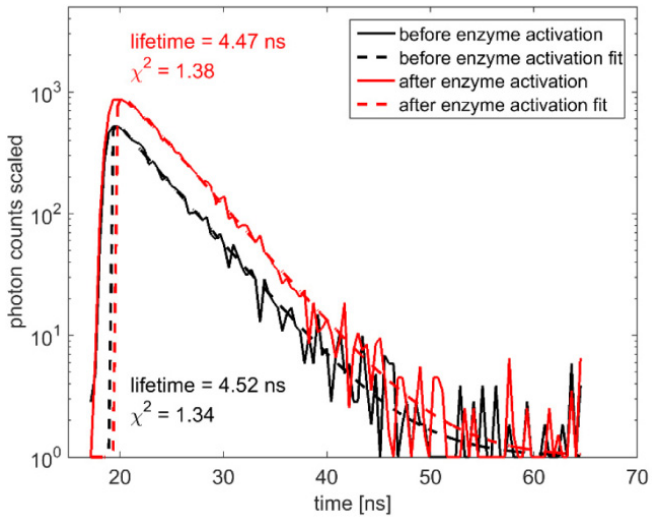

Fig. 16. Decays extracted from the fluorescence peak before enzyme activation (2 s time point) and after enzyme activation (130 s time point). 20 pixels were binned for both decays covering 520-528 nm.

3.4 TRFS of ex vivo lung

TRFS of *ex vivo* lung slide acquired in 258 ms is shown in Fig. 17. 10 pixels were binned for the TRFS in Fig. 17 resulting in 6 nm spectral resolution. The TRFS acquisition speed allows for point based analysis *in vivo*. For display purposes, etaloning fringes were removed in the same way as with FRET biosensor. We detect sharp broadband spectra at 15 ns in Fig. 17, which requires further investigation. It does not affect the spectral decays studied. The reduction in lifetime from 3.36 ns to 2.12 ns shown on Fig. 18(a) is due to presence of 1% Inspeck beads. This is also shown in lifetime change with wavelength on Fig. 18(b), top curve. Reduced χ^2 is above 3 across the spectrum (Fig. 18(b), bottom curve) indicating multi-exponential nature of the decay.

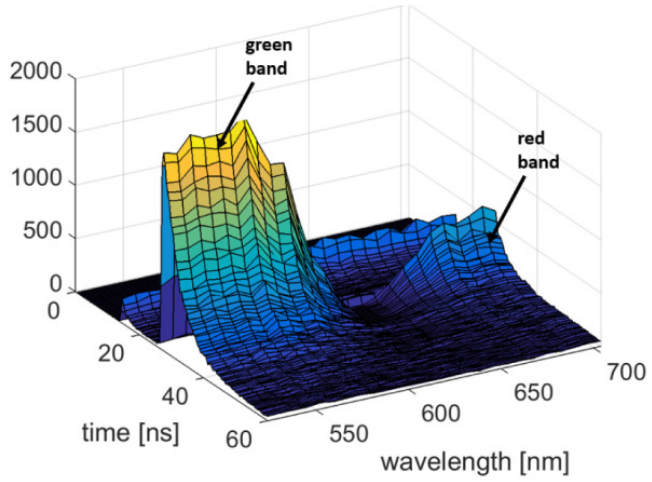

Fig. 17. TRFS of *ex vivo* lung acquired in 258 ms. 10 pixels were binned to obtain spectral resolution of 6 nm. Autofluorescence did not vary over 50 s measurements so no transient was observable.

Optics EXPRESS

However, if we use CMM for lifetime estimate, then spectral CMM can be acquired faster. The total number of photon events detected in 10 µs is ~1000 for the band 500 nm to 550 nm (CMM lung slide experiment data not shown). This indicates that a single binned CMM value for the whole green band is obtainable in 2 µs assuming CMM value is correct for 200 photon events. This is relevant when considering photon budgets required for laser scanning confocal FLIM, because it is indicative of 500000 lifetime estimates per second which is equivalent ~8 frames per second with 256 × 256 frames. The mouse lung slide used did not contain any fluorescent dye and it was 5 µm thick, so *in vivo* tissue imaging is likely to produce more autofluorescence. Furthermore, by deploying molecular imaging smartprobes [54,55] across the visible-NIR spectrum will allow us to use the CMOS SPAD line array for fast FLIM applications.

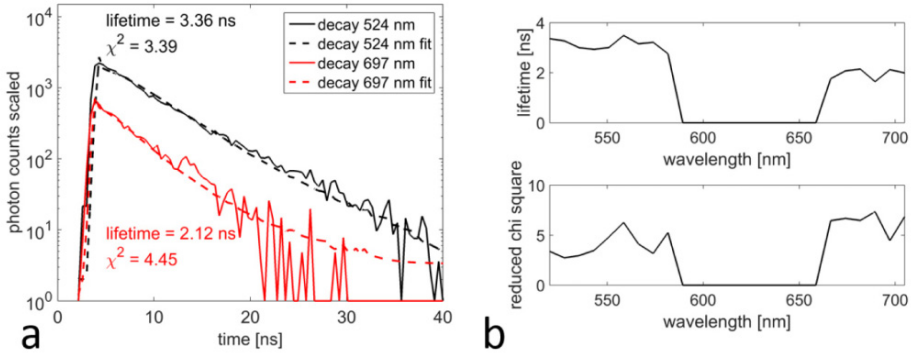

Fig. 18. Decays (from *ex vivo* lung tissue) with fits for green band (524 nm) and red band (697 nm) (a). Lifetime reduction in red band is shown in (b), top curve. Lifetime reduction is due to red Inspeck beads. Spectral resolution of each decay was 6 nm.

3.5 Time-gated Raman spectroscopy of toluene

Figure 19 demonstrates the time-gating performance of the CMOS SPAD line sensor. Time-gating was done in SPC mode. Exposure time was 5 s.The spectrum shown is from toluene focusing on expected Raman peak at 3056 cm^{-1}. Due to LRRS regime (linewidth of the laser is ~250 cm^{-1}) the peak is broad, but this still allows us to study the time-gating circuit performance. The background noise is suppressed by a factor of 3.6 by reducing the window for which the detector is on, thus reducing DCR. This is close to expected value of 3.57 (20 ns / 5.6 ns). Further demonstration of the application of this detector for removal of fluorescent background signals (in addition to DCR) has been shown elsewhere [13].

The background noise for both curves was estimated as standard deviation of values between 2000 cm^{-1} and 2500 cm^{-1} which are expected to be non-fluorescent. The main component of the background noise is DCR. However, besides DCR suppression, we can also observe that the main 3056 cm^{-1} signal peak is reduced from 3600 without time-gating to 1905 photon counts with time-gating. This needs further investigation and testing, but this is at least partially due to the time-gate missing a part of the diffuse tail of the red SPADs (see Fig. 8(b)).

Optics EXPRESS

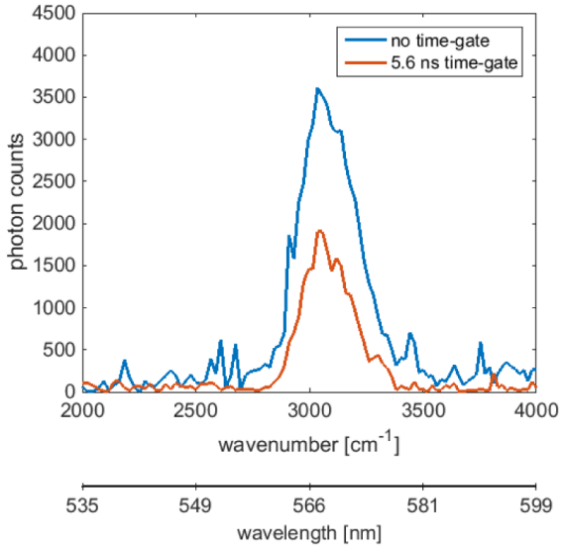

Fig. 19. Plot of non-time-gated (blue) spectrum and 5.6 ns time-gated Raman spectrum (red) of toluene. Exposure time was 5s and the acquisition was done in time-gated SPC mode.

4. Discussion

We demonstrated the versatility of the CMOS SPAD line array in several applications, but detailed work remains to be done. The main issues observed have been the coarse time resolution (0.43 ns), coarse jitter (at least 0.62 ns, but often > 1ns) and etaloning. Coarse timing limits lifetime analysis in the plant specimen where lifetime values are <1 ns. Better jitter and time resolution have been designed previously for time-resolved CMOS SPAD cameras [56], where time bin resolution was ~50 ps and the jitter was less than 200 ps. The aim for forthcoming CMOS SPAD line sensors [57] will be to introduce improved timing performance which will significantly enhance TCSPC TRFS and time-gating capability (e.g. DCR suppression). Etaloning was observed during lung slide and custom FRET probe experiments, but it was not observed during leaf experiments. This needs further investigation, as sensor induced etaloning expected from front illuminated light sensors should be observable in all specimens. Our current understanding is that the etaloning observed is induced by cuvette or slide.

Another important point is the multiplexing capability of pixels containing multiple SPADs (4 SPADs per pixel in our case), because this decides the minimum time required to obtain enough photons for lifetime estimation. While the number of photons required differs for single exponential and multi exponential fits, many applications can deal with lifetime derived from single exponential fits. Assuming 50 ns dead-time per SPAD, the number of photons that can be processed by a single TDC is 20 million photon events per second. This applies to TDC architectures with negligible conversion dead-time as is the case for gated ring oscillator architectures deployed in the CMOS SPAD line sensor presented. This fits nicely with using a 20 MHz pulsed laser. However, the situation changes if there are multiple SPADs per TDC per pixel. As soon as the TDC receives a STOP from the laser trigger (see Fig. 4) it is ready to be started by any of the remaining 3 SPADs while the SPAD used immediately prior to STOP is in dead-time. Therefore, in situations when there is a photon burst or samples emit many photons, the multiplexing allows more events to be processed. This is especially true if the repetition rate of the laser is increased. The key advantage brought by massively parallel CMOS SPAD timing circuits is that they can beat 1% pile-up limits by a wide margin [58] and exploit a variety of photon bursts to the full. The best results

Optics EXPRESS

we achieved are in chlorophyll A fluorescence transient analysis (section 3.2 above). We plan to expand this to spectral lifetime analysis of other fluorescence effects such as quenching and photobleaching and any transients observable in tissue autofluorescence and molecular imaging. A natural progression for our sensor is to implement it in confocal FLIM setup and demonstrate its performance in spectral imaging [11]. Frequency domain FLIM has already been deployed on plant cell FLIM at impressive speeds (26 fps) [59,60]. The advantage of massively parallelized CMOS SPAD sensors such as ours is that it can shift the study from measuring static fluorescence lifetime in either imaging or spectroscopy regime to measuring *rapid changes* in lifetime, intensity, spectra and other variables thus offering unique perspective in dynamic optical phenomena.

5. Conclusion

We demonstrate fast and flexible TRFS acquisition using TCSPC. Parallelization of TDC and time-gating speeds up the acquisition significantly for a range of applications and will enable advanced metrology such as spectral fluorescence kinetics in the ns and ms domains. Undoubtedly, 3D wafer stacking and process optimization [61] will enable faster, lower DCR and more integrated time-resolved sensors.

Funding

We would like to thank the Engineering and Physical Sciences Research Council (EPSRC, United Kingdom) Interdisciplinary Research Collaboration (grant number EP/K03197X/1) for funding this work. We would also like to thank EPSRC and MRC Centre for Doctoral Training in Optical Medical Imaging, OPTIMA, (grant number EP/L016559/1) for access to the supercontinuum laser. We would also like to thank ST Microelectronics, Imaging Division, Edinburgh, for their generous support in manufacturing of the CMOS SPAD line sensors.

Acknowledgments

We appreciate assistance from Dr Neil Finlayson, University of Edinburgh, with LRRS. Link to data sets can be found in [44].

Index

Bibliography

Abbas, T. A., Dutton, N. A. W., Almer, O., Pellegrini, S., Henrion, Y., and Henderson, R. K. Backside illuminated SPAD image sensor with 7.83μm pitch in 3D-stacked CMOS technology. In *2016 IEEE International Electron Devices Meeting (IEDM)*, pages 8.1.1–8.1.4, San Frasisco, CA, dec 2016. IEEE. ISBN 978-1-5090-3902-9 (electronic) 978-1-5090-3903-6 (print on demand). doi: 10.1109/IEDM.2016.7838372.

Agrawal, A., Pfefer, T. J., Woolliams, P. D., Tomlins, P. H., and Nehmetallah, G. Methods to assess sensitivity of optical coherence tomography systems. *Biomedical Optics Express*, 8(2):902, feb 2017. doi: 10.1364/boe.8.000902.

Akcay, C., Parrein, P., and Rolland, J. P. Estimation of longitudinal resolution in optical coherence imaging. *Applied Optics*, 41(25):5256, sep 2002. doi: 10.1364/AO.41.005256.

Al Abbas, T., Dutton, N. A., Almer, O., Finlayson, N., Rocca, F. M. D., and Henderson, R. A CMOS SPAD Sensor with a Multi-Event Folded Flash Time-to-Digital Converter for Ultra-Fast Optical Transient Capture. *IEEE Sensors Journal*, 18(8):3163–3173, apr 2018. doi: 10.1109/JSEN.2018.2803087.

Alfano, R. R. and Ockman, N. Methods for Detecting Weak Light Signals. *Journal of the Optical Society of America*, 58(1):90, jan 1968. doi: 10.1364/JOSA.58.000090.

An, L. and Wang, R. K. Use of a scanner to modulate spatial interferograms for in vivo full-range Fourier-domain optical coherence tomography. *Optics Letters*, 32 (23):3423–3425, dec 2007. doi: 10.1364/OL.32.003423.

An, L., Subhash, H. M., and Wang, R. K. Full range complex spectral domain optical coherence tomography for volumetric imaging at 47 000 A-scans per second. *Journal of Optics*, 12(8):084003, aug 2010. doi: 10.1088/2040-8978/12/8/084003.

Andor. Optical Etaloning in Charge Coupled Devices. URL `https://andor.oxinst.com/learning/view/article/optical-etaloning-in-charge-coupled-devices`. Accessed: 2019-02-06.

Andrews, L. C. *Special Functions of Mathematics for Engineers*. Oxford University Press, SPIE Press, 2nd edition, 1997. ISBN 9780819478467 (eBook) 9780819483713 (print).

Antolovic, I. M., Burri, S., Bruschini, C., Hoebe, R. A., and Charbon, E. SPAD imagers for super resolution localization microscopy enable analysis of fast fluorophore blinking. *Scientific Reports*, 7(March):44108, 2017. doi: 10.1038/srep44108.

Arlt, J., Tyndall, D., Rae, B. R., Li, D. D.-U., Richardson, J. A., and Henderson, R. K. A study of pile-up in integrated time-correlated single photon counting systems. *Review of Scientific Instruments*, 84:103105, oct 2013. doi: 10.1063/1.4824196.

Arzhantsev, S. and Maroncelli, M. Design and characterization of a femtosecond fluorescence spectrometer based on optical Kerr gating. *Applied Spectroscopy*, 59 (2):206–220, 2005. doi: 10.1366/0003702053085007.

Atkins, P., de Paula, J., and Smith, D. *Elements of Physical Chemistry*. Oxford University Press, 6th edition, 2013. ISBN 978-0-19-960811-9.

Ballew, R. M. and Demas, J. N. An Error Analysis of the Rapid Lifetime Determination Method for the Evaluation of Single Exponential Decays. *Analytical Chemistry*, 61 (1):30–33, 1989. doi: 10.1021/ac00176a007.

Barrick, J., Doblas, A., Gardner, M. R., Sears, P. R., Ostrowski, L. E., and Oldenburg, A. L. High-speed and high-sensitivity parallel spectral-domain optical coherence tomography using a supercontinuum light source. *Optics Letters*, 41(24):5620–5623, dec 2016. doi: 10.1364/ol.41.005620.

Baumann, B., Pircher, M., Götzinger, E., and Hitzenberger, C. K. Full range complex spectral domain optical coherence tomography without additional phase shifters. *Optics Express*, 15(20):13375–13387, oct 2007. doi: 10.1364/oe.15.013375.

Bec, J., Ma, D. M., Yankelevich, D. R., Liu, J., Ferrier, W. T., Southard, J., and Marcu, L. Multispectral fluorescence lifetime imaging system for intravascular diagnostics with ultrasound guidance: In vivo validation in swine arteries, may 2014. ISSN 1864-0648. URL `http://doi.wiley.com/10.1002/jbio.201200220`.

Becker, W. *Advanced Time-Correlated Single Photon Counting Techniques*. Springer-Verlag Berlin Heidelberg, 2005. ISBN 978-3-540-26047-9. doi: 10.1007/3-540-28882-1.

Becker, W. Fluorescence lifetime imaging by multi-dimensional time correlated single photon counting. *Medical Photonics*, 27:41–61, may 2015. doi: 10.1016/j.medpho.2015.02.001.

Becker, W., Bergmann, A., Biskup, C., Zimmer, T., Kloecker, N., and Benndorf, K. Multiwavelength TCSPC lifetime imaging. In Periasamy, A. and So, P. T. C., editors, *International Symposium on Biomedical Optics*, volume 4620, pages 79–84, San Jose, CA, jun 2002. SPIE. doi: 10.1117/12.470679.

Becker, W., Shcheslavkiy, V., Frere, S., and Slutsky, I. Spatially resolved recording of transient fluorescence-lifetime effects by line-scanning TCSPC. *Microscopy Research and Technique*, 77(3):216–224, jan 2014. doi: 10.1002/jemt.22331.

Becquerel, A.-E. Recherches sur les effets de la radiation chimique de la lumiere solaire au moyen des courants electriques. *Les Comptes Rendus de l'Académie des Sciences*, 9:145–149, 1839.

Beechem, J. M. Picosecond Fluorescence Decay Curves Collected on Millisecond Time Scale: Direct Measurement of Hydrodynamic Radii, Local/Global Mobility, and Intramolecular Distances during Protein-Folding Reactions. In Brand, L. and Johnson, M. L., editors, *Methods in Enzymology*, volume 278, chapter 3., pages 24–49. Academic Press, 1997. ISBN 978-0121821791. doi: 10.1016/S0076-6879(97)78005-4.

Behnel, S., Bradshaw, R., Citro, C., Dalcin, L., Seljebotn, D. S., and Smith, K. Cython: The best of both worlds. *Computing in Science and Engineering*, 13:31–39, mar 2011. doi: 10.1109/MCSE.2010.118.

Berezin, M. Y. and Achilefu, S. Fluorescence Lifetime Measurements and Biological Imaging. *Chemical Reviews*, 110(5):2641–2684, may 2010. doi: 10.1021/cr900343z.

Bezerra, H. G., Costa, M. A., Guagliumi, G., Rollins, A. M., and Simon, D. I. Intracoronary Optical Coherence Tomography: A Comprehensive Review. Clinical and Research Applications. *JACC: Cardiovascular Interventions*, 2(11):1035–1046, nov 2009. doi: 10.1016/j.jcin.2009.06.019.

Bradley, M., Chankeshwara, S. V., and Megia-Fernandez, A. Optical probe for thrombin, 2016. URL `https://patents.google.com/patent/WO2016151297A1/en`.

Braga, L. H., Pancheri, L., Gasparini, L., Perenzoni, M., Walker, R., Henderson, R. K., and Stoppa, D. A CMOS mini-SiPM detector with in-pixel data compression for PET applications. In *2011 IEEE Nuclear Science Symposium Conference Record*, pages 548–552. IEEE, oct 2011. ISBN 978-1-4673-0120-6 (electronic) 978-1-4673-0118-3 (print). doi: 10.1109/NSSMIC.2011.6154110.

Braga, L. H., Gasparini, L., Grant, L., Henderson, R. K., Massari, N., Perenzoni, M., Stoppa, D., and Walker, R. A fully digital 8x16 sipm array for pet applications with per-pixel tdcs and real-time energy output. *IEEE Journal of Solid-State Circuits*, 49 (1):301–314, 2014. doi: 10.1109/JSSC.2013.2284351.

Bronzi, D., Tisa, S., Villa, F., Bellisai, S., Tosi, A., and Zappa, F. Fast sensing and quenching of CMOS SPADs for minimal afterpulsing effects. *IEEE Photonics Technology Letters*, 25(8):776–779, 2013. doi: 10.1109/LPT.2013.2251621.

Bronzi, D., Villa, F., Tisa, S., Tosi, A., and Zappa, F. SPAD Figures of Merit for Photon-Counting, Photon-Timing, and Imaging Applications: A Review. *IEEE Sensors Journal*, 16(1):3–12, 2016. doi: 10.1109/JSEN.2015.2483565.

Bruschini, C., Homulle, H., Antolovic, I. M., Burri, S., and Charbon, E. Single-photon SPAD imagers in biophotonics: Review and Outlook. mar 2019.

Burns, D. T. Aspects of the development of colorimetric analysis and quantitative molecular spectroscopy in the ultraviolet-visible region. In Burgess, C. and Mielenz, K. D., editors, *Advances in Standards and Methodology in Spectrophotometry*, volume 2, pages 1–19. Elsevier, jan 1987. doi: 10.1016/B978-0-444-42880-6.50004-0.

Burns, D. T. Historical perspectives. Towards a definitive history of optical spectroscopy. Part II. Introduction of slits and collimator lens. Spectroscopes available before and just after Kirchoff and Bunsen{'}s studies. *Journal of Analytical Atomic Spectrometry*, 3(1):285–291, 1988. doi: 10.1039/JA9880300285.

Burri, S., Homulle, H., Bruschini, C., and Charbon, E. LinoSPAD: a time-resolved 256×1 CMOS SPAD line sensor system featuring 64 FPGA-based TDC channels running at up to 8.5 giga-events per second. In Berghmans, F. and Mignani, A. G., editors, *Proceedings Volume 9899, Optical Sensing and Detection IV*, volume 9899, page 98990D, Brussels, Belgium, 2016. SPIE. ISBN 9781510601444. doi: 10.1117/12.2227564.

Buttafava, M., Boso, G., Ruggeri, A., Dalla Mora, A., and Tosi, A. Time-gated single-photon detection module with 110 ps transition time and up to 80 MHz repetition rate. *Review of Scientific Instruments*, 85(8):083114, 2014. doi: 10.1063/1.4893385.

Butte, P. V., Fang, Q., Jo, J. A., Yong, W. H., Pikul, B. K., Black, K. L., and Marcu, L. Intraoperative delineation of primary brain tumors using time-resolved fluorescence spectroscopy. *Journal of Biomedical Optics*, 15(2):027008, 2010. doi: 10.1117/1.3374049.

Chang-Hasnain, C. J. Tunable VCSEL. *IEEE Journal on Selected Topics in Quantum Electronics*, 6(6):978–987, nov 2000. doi: 10.1109/2944.902146.

Charbon, E. Single-Photon imaging in complementary metal oxide semiconductor processes. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2012):20130100, mar 2014. doi: 10.1098/rsta.2013.0100.

Charbon, E., Gunther, N. J., Boiko, D. L., and Beretta, G. B. Design rules for quantum imaging devices: experimental progress using CMOS single-photon detectors. In Meyers, R. E., Shih, Y., and Deacon, K. S., editors, *Proceedings Volume 6305,*

Quantum Communications and Quantum Imaging IV, volume 6305, page 63050M, San Diego, CA, aug 2006. SPIE. doi: 10.1117/12.681235.

Chinn, S. R., Swanson, E. A., and Fujimoto, J. G. Optical coherence tomography using a frequency-tunable optical source. *Optics Letters*, 22(5):340–342, mar 1997. doi: 10.1364/OL.22.000340.

Chitnis, D. and Collins, S. A SPAD-based photon detecting system for optical communications. *Journal of Lightwave Technology*, 32(10):2028–2034, may 2014. doi: 10.1109/JLT.2014.2316972.

Choma, M., Sarunic, M., Yang, C., and Izatt, J. Sensitivity advantage of swept source and Fourier domain optical coherence tomography. *Optics Express*, 11(18):2183–2189, sep 2003. doi: 10.1364/OE.11.002183.

Clarke, R. H., Londhe, S., Premasiri, W. R., and Womble, M. E. Low-Resolution Raman Spectroscopy: Instrumentation and Applications in Chemical Analysis. *Journal of Raman Spectroscopy*, 30(9):827–832, sep 1999. doi: 10.1002/(SICI)1097-4555(199909)30:9<827::AID-JRS454>3.0.CO;2-S.

Cova, S., Ghioni, M., Lacaita, A., Samori, C., and Zappa, F. Avalanche photodiodes and quenching circuits for single-photon detection. *Applied Optics*, 35(12):1956–1976, 1996. doi: 10.1364/AO.35.001956.

Cova, S., Lacaita, A., and Ripamonti, G. Trapping Phenomena in Avalanche Photodiodes on Nanosecond Scale. *IEEE Electron Device Letters*, 12(12):685–687, 1991. doi: 10.1109/55.116955.

Day, P. K., LeDuc, H. G., Mazin, B. A., Vayonakis, A., and Zmuidzinas, J. A broadband superconducting detector suitable for use in large arrays. *Nature*, 425(6960):817–821, oct 2003. doi: 10.1038/nature02037.

de Boer, J. F., Cense, B., Park, B. H., Pierce, M. C., Tearney, G. J., and Bouma, B. E. Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. *Optics Letters*, 28(21):2067, nov 2003. doi: 10.1364/OL.28.002067.

de Boer, J. F., Leitgeb, R., and Wojtkowski, M. Twenty-five years of optical coherence tomography: the paradigm shift in sensitivity and speed provided by Fourier domain OCT [Invited]. *Biomedical Optics Express*, 8(7):3248, jul 2017. doi: 10.1364/BOE.8.003248.

del Valle, J. C. and Catalán, J. Kasha's rule: a reappraisal. *Physical Chemistry Chemical Physics*, 21(19):10061–10069, 2019. doi: 10.1039/C9CP00739C.

Dorrer, C., Belabas, N., Likforman, J.-P., and Joffre, M. Spectral resolution and sampling issues in Fourier-transform spectral interferometry. *Journal of the Optical Society of America B*, 17(10):1795, oct 2000. doi: 10.1364/josab.17.001795.

Drexler, W. and Fujimoto, J. G., editors. *Optical Coherence Tomography*. Springer International Publishing, 2 edition, 2015. ISBN 978-3-319-06418-5. doi: 10.1007/978-3-319-06419-2.

Drexler, W., Chen, Y., Aguirre, A. D., Považay, B., Unterhuber, A., and Fujimoto, J. G. Ultrahigh resolution optical coherence tomography. *Optical Coherence Tomography: Technology and Applications, Second Edition*, 9(1):277–318, 2015. doi: 10.1007/978-3-319-06419-2_9.

Dutton, N., Vergote, J., Gnecchi, S., Grant, L., Lee, D., Pellegrini, S., Rae, B., and Hendersony, R. Multiple-event direct to histogram TDC in 65nm FPGA technology. In *10th Conference on Ph. D. Research in Microelectronics and Electronics (PRIME)*, Grenoble, France, 2014. IEEE. ISBN 9781479949946. doi: 10.1109/prime.2014.6872727.

Efremov, E. V., Ariese, F., and Gooijer, C. Achievements in resonance Raman spectroscopy. Review of a technique with a distinct analytical chemistry potential. *Analytica Chimica Acta*, 606(2):119–134, jan 2008. doi: 10.1016/j.aca.2007.11.006.

Ehrlich, K., Kufcsák, A., Krstajić, N., Henderson, R. K., Thomson, R. R., and Tanner, M. G. Fibre optic time-resolved spectroscopy using CMOS-SPAD arrays. volume 10058, page 100580H. International Society for Optics and Photonics, feb 2017a. ISBN 9781510605572. doi: 10.1117/12.2252090.

Ehrlich, K., Kufcsák, A., McAughtrie, S., Fleming, H., Krstajic, N., Campbell, C. J., Henderson, R. K., Dhaliwal, K., Thomson, R. R., and Tanner, M. G. pH sensing through a single optical fibre using SERS and CMOS SPAD line arrays. *Optics Express*, 25(25):30976, dec 2017b. doi: 10.1364/OE.25.030976.

Ehrlich, K., Kufcsák, A., McAughtrie, S., Fleming, H., Krstajić, N., Campbell, C. J., Henderson, R. K., Dhaliwal, K., Thomson, R. R., and Tanner, M. G. Time-resolved single photon spectroscopy through a single optical fibre for miniaturised medical probe design. In Popp, J., Tuchin, V. V., and Pavone, F. S., editors, *Proceedings Volume 10685, Biophotonics: Photonic Solutions for Better Health Care VI*, volume 10685, page 24, Strasbourg, France, may 2018. SPIE. ISBN 9781510618961. doi: 10.1117/12.2307334.

Einstein, A. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, 322(6):132–148, 1905. doi: 10.1002/andp.19053220607.

Eisele, A., Henderson, R., Schmidtke, B., Funk, T., Grant, L., Richardson, J., and Freude, W. 185 MHz Count Rate, 139 dB Dynamic Range Single-Photon Avalanche Diode with Active Quenching Circuit in 130nm CMOS Technology. In *2011 International Image Sensor Workshop*, page R43, Hokkaido, Japan, 2011.

Elson, E. L. and Magde, D. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. *Biopolymers*, 13(1):1–27, jan 1974. doi: 10.1002/bip.1974.360130102.

Erdogan, A. T., Walker, R., Finlayson, N., Krstajić, N., Williams, G. O. S., and Henderson, R. K. A 16.5 Giga Events/s 1024 × 8 SPAD Line Sensor with per-pixel Zoomable 50ps-6.4ns/bin Histogramming TDC. In *2017 Symposium on VLSI Circuits*, pages C292 – C293, Kyoto, Japan, jun 2017. IEEE. ISBN 978-4-86348-614-0 (electronic) 978-1-5090-3980-7 (print on demand). doi: 10.23919/VLSIC.2017.8008513.

Erdogan, A. T., Walker, R., Finlayson, N., Krstajić, N., Williams, G. O. S., and Henderson, R. K. A CMOS SPAD Line Sensor with per-pixel Histogramming TDC for Time-Resolved Multispectral Imaging. *IEEE Journal of Solid-State Circuits*, in print:1–15, 2019. doi: 10.1109/JSSC.2019.2894355.

Ervin, J., Sabelko, J., and Gruebele, M. Submicrosecond real-time fluorescence sampling: Application to protein folding. *Journal of Photochemistry and Photobiology B: Biology*, 54(1):1–15, jan 2000. doi: 10.1016/S1011-1344(00)00002-6.

Evans, C. L. and Xie, X. S. Coherent Anti-Stokes Raman Scattering Microscopy: Chemical Imaging for Biology and Medicine. *Annual Review of Analytical Chemistry*, 1:883–909, jul 2008. doi: 10.1146/annurev.anchem.1.031207.112754.

Fang, Q., Papaioannou, T., Jo, J. A., Vaitha, R., Shastry, K., and Marcu, L. Time-domain laser-induced fluorescence spectroscopy apparatus for clinical diagnostics. *Review of Scientific Instruments*, 75(1):151–162, jan 2004. doi: 10.1063/1.1634354.

Fercher, A. F., Hitzenberger, C. K., Kamp, G., and El-Zaiat, S. Y. Measurement of intraocular distances by backscattering spectral interferometry. *Optics Communications*, 117(1-2):43–48, may 1995. doi: 10.1016/0030-4018(95)00119-S.

Fercher, A. F. Optical coherence tomography. *Journal of Biomedical Optics*, 1(2): 157–173, 1996. doi: 10.1117/12.231361.

Fercher, A. F., Leitgeb, R., Hitzenberger, C. K., Sattmann, H., and Wojtkowski, M. Complex spectral interferometry OCT. volume 3564, pages 173–178. International Society for Optics and Photonics, sep 1998. doi: 10.1117/12.339152.

Ferraro, John R., Nakamato, Kazuo Brown, C. W. *Introductory Raman Spectroscopy, Academic Press*. Academic Press, 2nd edition, 2003. ISBN 9780080509129 (eBook) 9780122541056 (print).

Feuda, R., Hamilton, S. C., McInerney, J. O., and Pisani, D. Metazoan opsin evolution reveals a simple route to animal vision. *Proceedings of the National Academy of Sciences*, 109(46):18868–18872, nov 2012. doi: 10.1073/pnas.1204609109.

Ficorella, A., Pancheri, L., Betta, G. F., Brogi, P., Collazuol, G., Marrocchesi, P. S., Morsani, F., Ratti, L., and Savoy-Navarro, A. Crosstalk mapping in CMOS SPAD arrays. In *2016 46th European Solid-State Device Research Conference (ESSDERC)*, volume 2016-Octob, pages 101–104, Lausanne, Switzerland, 2016. IEEE. ISBN 978-1-5090-2969-3 (electronic) 978-1-5090-2970-9 (print on demand). doi: 10.1109/ESSDERC.2016.7599598.

Finlayson, N., Erdogan, A. T., Usai, A., and Henderson, R. K. Noise characterization of a 512 x 16 spad line sensor for time-resolved spectroscopy applications. In Israel Gannot, editor, *Proceedings Volume 10488, Optical Fibers and Sensors for Medical Diagnostics and Treatment Applications XVIII*, volume 10488, page 15, San Frasisco, CA, 2018. SPIE. ISBN 9781510614611. doi: 10.1117/12.2290089.

Fishburn, M. W. *Fundamentals of CMOS single-photon avalanche diodes*. PhD thesis, Delft University of Technology, 2012.

Fossum, E. R. Active pixel sensors: are CCDs dinosaurs? In Blouke, M. M., editor, *Proceedings Volume 1900, Charge-Coupled Devices and Solid State Optical Sensors III*, volume 1900, pages 2–14, San Jose, CA, jul 1993. SPIE. doi: 10.1117/12.148585.

Fossum, E. R. The Quanta Image Sensor (QIS): Concepts and Challenges. In *Computational Optical Sensing and Imaging*, page JTuE1, Toronto, Canada, jul 2011. Optical Society of America. ISBN 978-1-55752-914-5. doi: 10.1364/COSI.2011.JTuE1.

Franck, F., Dewez, D., and Popovic, R. Changes in the Room-temperature Emission Spectrum of Chlorophyll During Fast and Slow Phases of the Kautsky Effect in Intact Leaves. *Photochemistry and Photobiology*, 81(2):431–436, apr 2005. doi: 10.1562/2004-03-01-RA-094.

Fraunhofer, J. Neue Modifikation des Lichtes durch gegenseitige Einwirkung und Beugung der Strahlen, und Gesetze derselben. In *Denkschriften der Königlichen Akademie der Wissenschaften zu München.*, pages 3–76. Königliche Akademie der Wissenschaften zu München, München :, 1821.

Fujimoto, J. G., Pitris, C., Boppart, S. A., and Brezinski, M. E. Optical Coherence Tomography: An Emerging Technology for Biomedical Imaging and Optical Biopsy. *Neoplasia*, 2(1-2):9–25, 2000. doi: doi:10.1038/sj.neo.7900071.

Gallivanoni, A., Rech, I., and Ghioni, M. Progress in quenching circuits for single photon avalanche diodes. *IEEE Transactions on Nuclear Science*, 57(6):3815–3826, 2010. doi: 10.1109/TNS.2010.2074213.

Gariepy, G., Krstajic, N., Henderson, R., Li, C., Thomson, R. R., Buller, G. S., Heshmat, B., Raskar, R., Leach, J., and Faccio, D. Single-photon sensitive light-in-fight imaging. *Nature Communications*, 6, 2015. doi: 10.1038/ncomms7021.

Gaviola, E. Ein Fluorometer. Apparat zur Messung von Fluoreszenzabklingungszeiten. *Zeitschrift für Physik*, 42(11-12):853–861, may 1927. doi: 10.1007/BF01776683.

Gersbach, M., Trimananda, R., Maruyama, Y., Fishburn, M., Stoppa, D., Richardson, J., Walker, R., Henderson, R. K., and Charbon, E. High frame-rate TCSPC-FLIM using a novel SPAD-based image sensor. In Dereniak, E. L., Hartke, J. P., LeVan, P. D., Sood, A. K., Longshore, R. E., and Razeghi, M., editors, *Proceedings Volume 7780, Detectors and Imaging Devices: Infrared, Focal Plane, Single Photon*, volume 7780, page 77801H, San Diego, CA, aug 2010. SPIE. ISBN 9780819482761. doi: 10.1117/12.860769.

Gnecchi, S., Dutton, N. A. W., Luca, P., Rae, B. R., Sara, P., McLeod, S. J., Grant, L. A., and Henderson, R. K. Digital Silicon Photomultipliers With OR/XOR Pulse Combining Techniques. *IEEE Transactions on Electron Devices*, 63(3):1105–1110, 2016. doi: 10.1109/TED.2016.2518301.

Gol'tsman, G. N., Okunev, O., Chulkova, G., Lipatov, A., Semenov, A., Smirnov, K., Voronov, B., Dzardanov, A., Williams, C., and Sobolewski, R. Picosecond superconducting single-photon optical detector. *Applied Physics Letters*, 79(6):705–707, 2001. doi: 10.1063/1.1388868.

Goodman, J. W. *Statistical Optics*. John Wiley & Sons, Inc., 2000. ISBN 0-47 1-0 1502-4.

Götzinger, E., Pircher, M., Leitgeb, R. A., and Hitzenberger, C. K. High speed full range complex spectral domain optical coherence tomography. *Optics Express*, 13 (2):583–594, jan 2005. doi: 10.1364/OPEX.13.000583.

Grajciar, B., Pircher, M., Hitzenberger, C. K., Findl, O., and Fercher, A. F. High sensitive measurement of the human axial eye length in vivo with Fourier domain low coherence interferometry. *Optics Express*, 16(4):2405–2414, feb 2008. doi: 10.1364/oe.16.002405.

Gu, M., Sheppard, C. J. R., and Gan, X. Image formation in a fiber-optical confocal scanning microscope. *Journal of the Optical Society of America A*, 8(11):1755–1761, nov 1991. doi: 10.1364/josaa.8.001755.

Gyongy, I., Calder, N., Davies, A., Dutton, N. A. W., Dalgarno, P., Duncan, R., Rickman, C., and Henderson, R. K. 256×256, 100kfps, 61% Fill-factor Time-resolved SPAD Image Sensor for Microscopy Applications. In *2016 IEEE International Electron Devices Meeting (IEDM)*, pages 8.2.1 – 8.2.4, San Frasisco, CA, 2016a. IEEE. ISBN 978-1-5090-3902-9 (electronic) 978-1-5090-3903-6 (print on demand). doi: 10.1109/IEDM.2016.7838373.

Gyongy, I., Davies, A., Dutton, N. A., Duncan, R. R., Rickman, C., Henderson, R. K., and Dalgarno, P. A. Smart-aggregation imaging for single molecule localisation with SPAD cameras. *Scientific Reports*, 6:1–10, 2016b. doi: 10.1038/srep37349.

Gyongy, I., Davies, A., Gallinet, B., Dutton, N. A., Duncan, R. R., Rickman, C., Henderson, R. K., and Dalgarno, P. A. Cylindrical microlensing for enhanced collection efficiency of small pixel SPAD arrays in single-molecule localisation microscopy. *Optics Express*, 26(3):2280–2291, feb 2018. doi: 10.1364/OE.26.002280.

Hamamatsu. *PHOTOMULTIPLIER TUBES*. Hamamatsu Photonics K. K., third edition, 2007.

Haupt, C. and Huber, A. B. How axons see their way - axonal guidance in the visual system. *Frontiers in Bioscience*, 13(1):3136–3149, 2008. doi: 10.2741/2915.

Häusler, G. and Lindner, M. W. "Coherence Radar" And "Spectral Radar"—New Tools for Dermatological Diagnosis. *Journal of Biomedical Optics*, 3(1):21–31, 1998. doi: 10.1117/1.429899.

Hearnshaw, J. B. *The analysis of starlight: two centuries of astronomical spectroscopy*. Cambridge University Press, 2nd edition, 2014. ISBN 978-1107031746. doi: 10.1017/CBO9781139382779.

Herschel, J. F. W. On the Absorption of Light by Coloured Media, and on the Colours of the Prismatic Spectrum exhibited by certain Flames; with an Account of a ready Mode of determining the absolute dispersive Power of any Medium, by direct experiment. *Transactions of the Royal Society of Edinburgh*, 9:445–460, 1823. doi: 10.1038/001534a0.

Herschel, W. Experiments on the refrangibility of the invisible rays of the sun. *Philosophical Transactions of the Royal Society of London*, 90:284–292, 1800. doi: 10.1080/14786440008562602.

Hertz, H. Ueber einen Einfluss des ultravioletten Lichtes auf die electrische Entladung. *Annalen der Physik*, 267(8):983–1000, 1887. doi: 10.1002/andp.18872670827.

Hine, H. M. Rome, the Cosmos, and the Emperor in Seneca ' s Natural Questions *. *The Journal of Roman Studies*, 96:42–72, 2006. doi: 10.3815/000000006784016224.

Hitzenberger, C. K., Baumgartner, A., Drexler, W., and Fercher, A. F. Dispersion Effects in Partial Coherence Interferometry: Implications for Intraocular Ranging. *Journal of Biomedical Optics*, 4(1):144–151, jan 1999. doi: 10.1117/1.429900.

Ho, J., Castro, D. P., Castro, L. C., Chen, Y., Liu, J., Mattox, C., Krishnan, C., Fujimoto, J. G., Schuman, J. S., and Duker, J. S. Clinical Assessment of Mirror Artifacts in Spectral-Domain Optical Coherence Tomography. *Investigative Ophthalmology and Visual Science*, 51(7):3714–3720, jul 2010. doi: 10.1167/iovs.09-4057.

Holmes, J. and Hattersley, S. Image blending and speckle noise reduction in multi-beam OCT. In Fujimoto, J. G., Izatt, J. A., and Tuchin, V. V., editors, *Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XIII*, volume 7168, page 71681N. SPIE, feb 2009. doi: 10.1117/12.808575.

Holmes, J., Hattersley, S., Stone, N., Bazant-Hegemark, F., and Barr, H. Multi-channel Fourier domain OCT system with superior lateral resolution for biomedical applications. In Izatt, J. A., Fujimoto, J. G., and Tuchin, V. V., editors, *Coherence Domain Optical Methods and Optical Coherence Tomography in Biomedicine XII*, volume 6847, page 68470O. SPIE, feb 2008. ISBN 9780819470225. doi: 10.1117/12.761655.

Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., Hee, M. R., Flotte, T., Gregory, K., Puliafito, C. A., and Fujimoto, J. G. Optical Coherence Tomography. *Science*, 254(5035):1178–1181, 1991. doi: 10.1126/science.1957169.

Huang, Y. and Kang, J. U. Real-time reference A-line subtraction rate Fourier-domain optical coherence graphics processing unit for high-frame- and saturation artifact removal using tomography video imaging. *Optical Engineering*, 51(7):073203, jul 2012. doi: 10.1117/1.OE.51.7.073203.

Huber, R., Wojtkowski, M., Taira, K., Fujimoto, J. G., and Hsu, K. Amplified, frequency swept lasers for frequency domain reflectometry and OCT imaging: design and scaling principles. *Optics Express*, 13(9):3513–3528, may 2005. doi: 10.1364/OPEX.13.003513.

Huber, R., Wojtkowski, M., and Fujimoto, J. G. Fourier Domain Mode Locking (FDML): A new laser operating regime and applications for optical coherence tomography. *Optics Express*, 14(8):3225–3237, apr 2006. doi: 10.1364/OE.14.003225.

Huggins, W. and Miller, W. A. On the Spectra of Some of the Fixed Stars. *Philosophical Transactions of the Royal Society of London*, 154:413–435, 1864. doi: 10.1098/rstl.1864.0012.

Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

Hynecek, J. Impactron—A New Solid State Image Intensifier. *IEEE Transactions on Electron Devices*, 48(10):2238–2241, 2001. doi: 10.1109/16.954460.

Iams, H. and Salzberg, B. The Secondary Emission Phototube. *Proceedings of the Institute of Radio Engineers*, 23(1):55–64, jan 1935. doi: 10.1109/JRPROC.1935.227243.

Irwin, K. D. and Hilton, G. C. Transition-Edge Sensors. In Enss, C., editor, *Cryogenic Particle Detection*, pages 63–150. Springer Berlin Heidelberg, 2005. ISBN 978-3-540-20113-7 (print) 978-3-540-31478-3 (online). doi: 10.1007/10933596_3.

IS-Instruments. Spectral Response Correction: The Effect of Etaloning, 2017. URL `https://www.azom.com/article.aspx?ArticleID=14009`. Accessed: 2019-02-06.

Ishay, E. B., Hazan, G., Rahamim, G., Amir, D., and Haas, E. An instrument for fast acquisition of fluorescence decay curves at picosecond resolution designed for double kinetics experiments: Application to fluorescence resonance excitation energy transfer study of protein folding. *Review of Scientific Instruments*, 83(8):084301, aug 2012. doi: 10.1063/1.4737632.

Jacques, S. L. Optical properties of biological tissues: A review, 2013. ISSN 1361-6560 (online) 0031-9155 (print). URL `http://iopscience.iop.org/0031-9155/58/11/R37`.

Jares-Erijman, E. A. and Jovin, T. M. FRET imaging. *Nature Biotechnology*, 21(11):1387–1395, nov 2003. doi: 10.1038/nbt896.

Jungwirth, J., Baumann, B., Pircher, M., Götzinger, E., and Hitzenberger, C. K. Extended in vivo anterior eye-segment imaging with full-range complex spectral domain optical coherence tomography. *Journal of Biomedical Optics*, 14(5):050501, 2009. doi: 10.1117/1.3213569.

Kalaji, H. M., Schansker, G., Ladle, R. J., Goltsev, V., Bosa, K., Allakhverdiev, S. I., Brestic, M., Bussotti, F., Calatayud, A., Dąbrowski, P., Elsheery, N. I., Ferroni, L., Guidi, L., Hogewoning, S. W., Jajoo, A., Misra, A. N., Nebauer, S. G., Pancaldi, S., Penella, C., Poli, D., Pollastrini, M., Romanowska-Duda, Z. B., Rutkowska, B., Serôdio, J., Suresh, K., Szulc, W., Tambussi, E., Yanniccari, M., and Zivcak, M. Frequently asked questions about in vivo chlorophyll fluorescence: practical issues. *Photosynthesis Research*, 122(2):121–158, 2014. doi: 10.1007/s11120-014-0024-6.

Kalisz, J. Review of methods for time interval measurements with picosecond resolution. *Metrologia*, 41(1):17–32, 2004. doi: 10.1088/0026-1394/41/1/004.

Kalkman, J. Fourier-Domain Optical Coherence Tomography Signal Analysis and Numerical Modeling. *International Journal of Optics*, 2017:1–16, mar 2017. doi: 10.1155/2017/9586067.

Kathuria, S. V. and Bilsel, O. *Probing Microsecond Reactions with Microfluidic Mixers and TCSPC*, pages 357–384. Springer International Publishing, Cham, 2015. ISBN 978-3-319-14928-8 (print) 978-3-319-14929-5 (online). doi: 10.1007/978-3-319-14929-5_11.

Kekkonen, J., Nissinen, J., Kostamovaara, J., and Nissinen, I. Distance-Resolving Raman Radar Based on a Time-Correlated CMOS Single-Photon Avalanche Diode Line Sensor. *Sensors*, 18(10):3200, sep 2018. doi: 10.3390/s18103200.

Kim, J.-h., Han, J.-H., and Jeong, J. Adaptive optimization of reference intensity for optical coherence imaging using galvanometric mirror tilting method. *Optics Communications*, 351:57–62, sep 2015. doi: 10.1016/j.optcom.2015.04.041.

Kirchhoff, G. I. On the relation between the radiating and absorbing powers of different bodies for light and heat. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 20(130):1–21, 1860. doi: 10.1080/14786446008642901.

Kiwanuka, S.-S., Laurila, T. K., Frank, J. H., Esposito, A., Blomberg Von Der Geest, K., Pancheri, L., Stoppa, D., and Kaminski, C. F. Development of Broadband Cavity Ring-Down Spectroscopy for Biomedical Diagnostics of Liquid Analytes. *Analytical Chemistry*, 84(13):5489–5493, jul 2012. doi: 10.1021/ac301108q.

Kodach, V. M., Kalkman, J., Faber, D. J., and van Leeuwen, T. G. Quantitative comparison of the OCT imaging depth at 1300 nm and 1600 nm. *Biomedical Optics Express*, 1(1):176–185, jul 2010. doi: 10.1364/BOE.1.000176.

Köllner, M. and Wolfrum, J. How many photons are necessary for fluorescence-lifetime measurements? *Chemical Physics Letters*, 200(1-2):199–204, nov 1992. doi: 10.1016/0009-2614(92)87068-Z.

Kragh, H. The names of physics: plasma, fission, photon. *European Physical Journal H*, 39(3):263–281, sep 2014. doi: 10.1140/epjh/e2014-50007-7.

Krstajić, N., Brown, C. T. A., Dholakia, K., and Giardini, M. E. Tissue surface as the reference arm in Fourier domain optical coherence tomography. *Journal of Biomedical Optics*, 17(7):071305, may 2012. doi: 10.1117/1.JBO.17.7.071305.

Krstajić, N., Levitt, J., Poland, S., Ameer-Beg, S., and Henderson, R. 256 × 2 SPAD line sensor for time resolved fluorescence spectroscopy. *Optics Express*, 23(5): 5653–5669, 2015. doi: 10.1364/OE.23.005653.

Krstajic, N., Poland, S., Levitt, J., Walker, R., Erdogan, A., Ameer-Beg, S., and Henderson, R. K. 0.5 billion events per second time correlated single photon counting using CMOS SPAD arrays. *Optics Letters*, 40(18):4305–4308, sep 2015. doi: 10.1364/OL.40.004305.

Kubetsky, L. A. Multiple amplifier. *Proceedings of the Institute of Radio Engineers*, 25 (4):421–433, apr 1937. doi: 10.1109/JRPROC.1937.229045.

Kufcsák, A., Erdogan, A., Walker, R., Ehrlich, K., Tanner, M., Megia-Fernandez, A., Scholefield, E., Emanuel, P., Dhaliwal, K., Bradley, M., Henderson, R. K., and Krstajić, N. Time-resolved spectroscopy at 19,000 lines per second using a CMOS SPAD line array enables advanced biophotonics applications. *Optics Express*, 25 (10):11103–11123, may 2017. doi: 10.1364/OE.25.011103.

Kukura, P., McCamant, D. W., and Mathies, R. A. Femtosecond Stimulated Raman Spectroscopy. *Annual Review of Physical Chemistry*, 58:461–488, may 2007. doi: 10.1146/annurev.physchem.58.032806.104456.

Kuo, W.-C., Lai, C.-M., Huang, Y.-S., Chang, C.-Y., and Kuo, Y.-M. Balanced detection for spectral domain optical coherence tomography. *Optics Express*, 21(16):19280, aug 2013. doi: 10.1364/oe.21.019280.

Labruyère, A., Tonello, A., Couderc, V., Huss, G., and Leproux, P. Compact supercontinuum sources and their biomedical applications. *Optical Fiber Technology*, 18(5): 375–378, sep 2012. doi: 10.1016/j.yofte.2012.08.003.

Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. Springer, 3rd edition, 2006. ISBN 978-0 387-31278-1 (print) 978-0-387-46312-4 (eBook). doi: 10.1007/978-0-387-46312-4.

LaRocca, F., Chiu, S. J., McNabb, R. P., Kuo, A. N., Izatt, J. A., and Farsiu, S. Robust automatic segmentation of corneal layer boundaries in SDOCT images using graph theory and dynamic programming. *Biomedical Optics Express*, 2(6):1524–1538, jun 2011. doi: 10.1364/BOE.2.001524.

Lee, C.-K., Tsai, M.-T., and Wu, C.-T. A pseudo-spectrum reconstruction method for reducing saturation artifact in spectral-domain optical coherence tomography. In Popp, J., Tuchin, V. V., Matthews, D. L., Pavone, F. S., and Garside, P., editors, *Proceedings Volume 9129, Biophotonics: Photonic Solutions for Better Health Care IV*, volume 9129, page 91290K. SPIE, may 2014. ISBN 9781628410778. doi: 10.1117/12.2054721.

Lee, S. H. and Gardner, R. P. A new G-M counter dead time model. *Applied Radiation and Isotopes*, 53(4-5):731–737, 2000. doi: 10.1016/S0969-8043(00)00261-X.

Leitgeb, R., Hitzenberger, C. K., and Fercher, A. F. Performance of fourier domain vs. time domain optical coherence tomography. *Optics Express*, 11(8):889–894, apr 2003. doi: 10.1364/OE.11.000889.

Leitgeb, R. A., Villiger, M., Bachmann, A. H., Steinmann, L., and Lasser, T. Extended focus depth for Fourier domain optical coherence microscopy. *Optics Letters*, 31 (16):2450–2452, aug 2006. doi: 10.1364/OL.31.002450.

Leitgeb, R. A. and Baumann, B. Multimodal Optical Medical Imaging Concepts Based on Optical Coherence Tomography. *Frontiers in Physics*, 6:114, oct 2018. doi: 10.3389/fphy.2018.00114.

Leitgeb, R. A., Michaely, R., Lasser, T., and Sekhar, S. C. Complex ambiguity-free Fourier domain optical coherence tomography through transverse scanning. *Optics Letters*, 32(23):3453–3455, dec 2007. doi: 10.1364/ol.32.003453.

Lexer, F., Hitzenberger, C. K., Drexler, W., Molebny, S., Sattmann, H., Sticker, M., and Fercher, A. F. Dynamic coherent focus OCT with depth- independent transversal resolution. *Journal of Modern Optics*, 46(3):541–553, 1999. doi: 10.1080/09500349908231282.

Li, C., Wang, J., Kobayashi, H., and Shiota, R. Timing measurement BOST architecture with full digital circuit and self-calibration using characteristics variation positively for fine time resolution. In *2016 IEEE 21st International Mixed-Signal Testing Workshop, IMSTW 2016*. Institute of Electrical and Electronics Engineers Inc., jul 2016. ISBN 9781509027514. doi: 10.1109/IMS3TW.2016.7524228.

Li, D. D.-U., Arlt, J., Tyndall, D., Walker, R., Richardson, J., Stoppa, D., Charbon, E., and Henderson, R. K. Video-rate fluorescence lifetime imaging camera with CMOS single-photon avalanche diode arrays and high-speed imaging algorithm. *Journal of Biomedical Optics*, 16(9):096012, 2011. doi: 10.1117/1.3625288.

Li, D. D.-U., Ameer-Beg, S., Arlt, J., Tyndall, D., Walker, R., Matthews, D. R., Visitkul, V., Richardson, J., and Henderson, R. K. Time-Domain Fluorescence Lifetime Imaging Techniques Suitable for Solid-State Imaging Sensor Arrays. *Sensors*, 12(5):5650–5669, may 2012. doi: 10.3390/s120505650.

Li, D.-U., Walker, R., Richardson, J., Rae, B., Buts, A., Renshaw, D., and Henderson, R. Hardware implementation and calibration of background noise for an integration-based fluorescence lifetime sensing algorithm. *Journal of the Optical Society of America A*, 26(4):804–14, 2009. doi: 10.1364/JOSAA.26.000804.

Li, D.-U., Rae, B., Andrews, R., Arlt, J., and Henderson, R. K. Hardware implementation algorithm and error analysis of high-speed fluorescence lifetime sensing systems using center-of-mass method. *Journal of Biomedical Optics*, 15(1):017006, jan 2010. doi: 10.1117/1.3428872.

Li, Z. and Deen, M. J. Towards a portable Raman spectrometer using a concave grating and a time-gated CMOS SPAD. *Optics Express*, 22(15):18736–18747, 2014. doi: 10.1364/OE.22.018736.

Lloyd, W. R., Wilson, R. H., Chang, C.-W., Gillispie, G. D., and Mycek, M.-A. Instrumentation to rapidly acquire fluorescence wavelength-time matrices of biological tissues. *Biomedical Optics Express*, 1(2):574 – 586, aug 2010. doi: 10.1364/BOE.1.000574.

Ma, J., Masoodian, S., Starkey, D. A., and Fossum, E. R. Photon-number-resolving megapixel image sensor at room temperature without avalanche gain. *Optica*, 4 (12):1474–1481, dec 2017. doi: 10.1364/OPTICA.4.0014741.

Maarek, J.-M. I., Marcu, L., Fishbein, M. C., and Grundfest, W. S. Time-Resolved Fluorescence of Human Aortic Wall: Use for Improved Identification of Atherosclerotic Lesions. *Lasers in Surgery and Medicine*, 27(3):241–254, jan 2000. doi: 10.1002/1096-9101(2000)27:3<241::AID-LSM6>3.0.CO;2-0.

Mandai, S. and Charbon, E. Multi-Channel Digital SiPMs: Concept, Analysis and Implementation. In *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, pages 1840–1844. IEEE, oct 2012. ISBN 978-1-4673-2030-6 (electronic) 978-1-4673-2028-3 (print). doi: 10.1109/NSSMIC.2012.6551429.

Marcu, L. Fluorescence Lifetime Techniques in Medical Applications. *Annals of Biomedical Engineering*, 40(2):304–331, feb 2012. doi: 10.1007/s10439-011-0495-y.

Marcu, L., Grundfest, W. S., and Maarek, J. M. I. Photobleaching of Arterial Fluorescent Compounds: Characterization of Elastin, Collagen and Cholesterol Time-resolved Spectra during Prolonged Ultraviolet Irradiation Laura. *Photochemistry and Photobiology*, 69(6):713–721, jun 1999. doi: 10.1111/j.1751-1097.1999.tb03352.x.

Marcu, L., French, P. M. W., and Elson, D. S. *Fluorescence Lifetime Spectroscopy and Imaging: Principles and Applications in Biomedical Diagnostics*. CRC Press, 1st edition, 2014. ISBN 9781439861677.

Marsili, F., Verma, V. B., Stern, J. A., Harrington, S., Lita, A. E., Gerrits, T., Vayshenker, I., Baek, B., Shaw, M. D., Mirin, R. P., and Nam, S. W. Detecting single infrared photons with 93% system efficiency. *Nature Photonics*, 7:210–214, mar 2013. doi: 10.1038/NPHOTON.2013.13.

Maruyama, Y. and Charbon, E. An all-digital, time-gated 128X128 spad array for on-chip, filter-less fluorescence detection. In *2011 16th International Solid-State Sensors, Actuators and Microsystems Conference*, pages 1180–1183, Beijing, China, 2011. IEEE. ISBN 978-1-4577-0156-6 (electronic) 978-1-4577-0157-3 (print). doi: 10.1109/TRANSDUCERS.2011.5969324.

Maruyama, Y., Blacksberg, J., and Charbon, E. A 1024×8, 700-ps Time-Gated SPAD Line Sensor for Planetary Surface Exploration With Laser Raman Spectroscopy and LIBS. *IEEE Journal of Solid-State Circuits*, 49(1):179–189, 2014. doi: 10.1109/JSSC.2013.2282091.

Maxwell, K. and Johnson, G. N. Chlorophyll fluorescence—a practical guide. *Journal of Experimental Botany*, 51(345):659–668, apr 2000. doi: 10.1093/jexbot/51.345.659.

McCreery, R. L. *Raman Spectroscopy for Chemical Analysis*. John Wiley & Sons, Inc., 2000. ISBN 9780471252870 (print) 9780471721642 (online). doi: 10.1002/0471721646.

Menninga, H., Favi, C., Fishburn, M. W., and Charbon, E. A Multi-Channel, lOps Resolution, FPGA-Based TDC with 300MS/s Throughput for Open-Source PET Applications Harmen. In *2011 IEEE Nuclear Science Symposium Conference Record*, pages 1515–1522, Valencia, Spain, 2012. IEEE. ISBN 978-1-4673-0120-6 (electonic) 978-1-4673-0118-3 (print). doi: 10.1109/NSSMIC.2011.6154362.

Michalet, X., Weiss, S., and Jäger, M. Single-Molecule Fluorescence Studies of Protein Folding and Conformational Dynamics. *Chemical Reviews*, 106(5):1785–1813, 2006. doi: 10.1021/cr0404343.

Millar, D. P. Time resolved fluorescence spectroscopy. *Current Opinion in Structural Biology*, 6(5):637–642, oct 1996. doi: 10.1007/BF02842096.

Mills, B., Bradley, M., and Dhaliwal, K. Optical imaging of bacterial infections. *Clinical and Translational Imaging*, 4(3):163–174, jun 2016. doi: 10.1007/s40336-016-0180-0.

Moerner, W. E. and Fromm, D. P. Methods of single-molecule fluorescence spectroscopy and microscopy. *Review of Scientific Instruments*, 74(8):3597–3619, aug 2003. doi: 10.1063/1.1589587.

Mohan, N., Minaeva, O., Gol'tsman, G. N., Nasr, M. B., Saleh, B. E., Sergienko, A. V., and Teich, M. C. Photon-counting optical coherence-domain reflectometry using superconducting single- photon detectors. *Optics Express*, 16(22):18118–18130, oct 2008. doi: 10.1364/OE.16.018118.

Movasaghi, Z., Rehman, S., and ur Rehman, D. I. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews*, 43(2):134–179, 2008. doi: 10.1080/05704920701829043.

Muller, M. S. and Fraser, J. M. Contrast improvement in Fourier-domain optical coherence tomography through time gating. *Journal of the Optical Society of America A*, 26(4):969–976, apr 2009. doi: 10.1364/JOSAA.26.000969.

Natarajan, C. M., Tanner, M. G., and Hadfield, R. H. Superconducting nanowire single-photon detectors: physics and applications. *Superconductor Science and Technology*, 25(6):063001, 2012. doi: 10.1088/0953-2048/25/6/063001.

Newton, I. A Letter of Mr. Isaac Newton, Professor of the Mathematicks in the University of Cambridge; Containing His New Theory about Light and Colors: Sent by the Author to the Publisher from Cambridge, Febr. 6. 1671/72; In Order to be Communicated to the R. Socie. *Philosophical Transactions of the Royal Society of London*, 6(69-80):3075–3087, jan 1671. doi: 10.1098/rstl.1671.0072.

Niclass, C., Rochas, A., Besse, P.-A., and Charbon, E. Design and Characterization of a CMOS 3-D Image Sensor Based on Single Photon Avalanche Diodes. *IEEE Journal of Solid-State Circuits*, 40(9):1847–1854, 2005. doi: 10.1109/JSSC.2005.848173.

Nissinen, I., Nissinen, J., Länsman, A.-K., Hallman, L., Kilpelä, A., Kostamovaara, J., Kögler, M., Aikio, M., and Tenhunen, J. A Sub-ns Time-gated CMOS Single Photon Avalanche Diode Detector for Raman Spectroscopy. In *2011 Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, pages 375–378, Helsinki, Finland, 2011. IEEE. ISBN 978-1-4577-0708-7 (electronic) 978-1-4577-0707-0 (print). doi: 10.1109/ESSDERC.2011.6044156.

Nissinen, I., Nissinen, J., Keränen, P., Stoppa, D., and Kostamovaara, J. A 16x256 SPAD Line Detector with a 50-ps, 3- bit, 256-channel Time-to-Digital Converter for Raman Spectroscopy. *IEEE Sensors Journal*, 18(9):3789–3798, 2018. doi: 10.1109/JSEN.2018.2813531.

Oliphant, T. E. *A guide to NumPy*. CreateSpace Independent Publishing Platform, 2nd edition, 2015. ISBN 9781517300074.

Palchetti, L., Bianchini, G., Cortesi, U., Pascale, E., and Clare, L. Assessment of Detector Nonlinearity in Fourier Transform Spectroscopy. *Applied Spectroscopy*, 56 (2):271–274, 2002. doi: 10.1366/0003702021954548.

Palubiak, D. P. and Deen, M. J. CMOS SPADs: Design Issues and Research Challenges for Detectors, Circuits, and Arrays. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(6):6000718, 2014. doi: 10.1109/JSTQE.2014.2344034.

Palubiak, D. P., Li, Z., and Deen, M. J. Afterpulsing Characteristics of Free-Running and Time-Gated Single-Photon Avalanche Diodes in 130-nm CMOS. *IEEE Transactions on Electron Devices*, 62(11):3727–3733, 2015. doi: 10.1109/TED.2015.2475126.

Pancheri, L., Massari, N., and Stoppa, D. SPAD Image Sensor With Analog Counting Pixel for Time-Resolved Fluorescence Detection. *IEEE Transactions on Electron Devices*, 60(10):3442–3449, oct 2013. doi: 10.1109/TED.2013.2276752.

Para, A. Afterpulsing in Silicon Photomultipliers: Impact on the Photodetectors Characterization, 2015.

Pavia, J. M., Wolf, M., and Charbon, E. Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery. *Optics Express*, 22(4):4202–4213, feb 2014. doi: 10.1364/OE.22.004202.

Pavia, J. M., Scandini, M., Lindner, S., Wolf, M., and Charbon, E. A 1 x 400 Backside-Illuminated SPAD Sensor with 49.7 ps Resolution, 30 pJ/Sample TDCs Fabricated in 3D CMOS Technology for Near-Infrared Optical Tomography. *IEEE Journal of Solid-State Circuits*, 50(10):2406–2418, 2015. doi: 10.1109/JSSC.2015.2467170.

Peacock, A., Verhoeve, P., Rando, N., Van Dordrecht, A., Taylor, B. G., Erd, C., Perryman, M. A., Venn, R., Howlett, J., Goldie, D. J., Lumley, J., and Wallis, M. Single optical photon detection with a superconducting tunnel junction. *Nature*, 381:135–137, may 1996. doi: 10.1038/381135a0.

Pellegrini, S. and Rae, B. Fully industrialised single photon avalanche diodes. In Itzler, M. A. and Campbell, J. C., editors, *Proceedings Volume 10212, Advanced Photon Counting Techniques XI*, volume 10212, page 102120D, Anaheim, CA, may 2017. SPIE. ISBN 9781510609259. doi: 10.1117/12.2264364.

Pellegrini, S., Warburton, R. E., Tan, L. J., Ng, J. S., Krysa, A. B., Groom, K., David, J. P. R., Cova, S., Robertson, M. J., and Buller, G. S. Design and Performance of an InGaAs–InP Single-Photon Avalanche Diode Detector. *IEEE Journal of Quantum Electronics*, 42(4):397–403, 2006. doi: 10.1109/jqe.2006.871067.

Pfefer, T. J., Paithankar, D. Y., Poneros, J. M., Schomacker, K. T., and Nishioka, N. S. Temporally and Spectrally Resolved Fluorescence Spectroscopy for the Detection of High Grade Dysplasia in Barrett's Esophagus. *Lasers in Surgery and Medicine*, 32(1):10–16, 2003. doi: 10.1002/lsm.10136.

PicoQuant. Fluorescence Microscope. URL `https://www.picoquant.com/applications/category/materials-science/time-resolved-photoluminescence{#}images`.

Pitts, J. D. and Mycek, M.-A. Design and development of a rapid acquisition laser-based fluorometer with simultaneous spectral and temporal resolution. *Review of Scientific Instruments*, 72(7):3061–3072, jul 2001. doi: 10.1063/1.1379957.

Poland, S. P., Krstajić, N., Monypenny, J., Coelho, S., Tyndall, D., Walker, R. J., Devauges, V., Richardson, J., Dutton, N., Barber, P., Li, D. D.-U., Suhling, K., Ng, T., Henderson, R. K., and Ameer-Beg, S. M. A high speed multifocal multiphoton fluorescence lifetime imaging microscope for live-cell FRET imaging. *Biomedical Optics Express*, 6(2):277–296, feb 2015. doi: 10.1364/BOE.6.000277.

Poland, S. P., Erdogan, A. T., Krstajić, N., Levitt, J., Devauges, V., Walker, R. J., Li, D. D.-U., Ameer-Beg, S. M., and Henderson, R. K. New high-speed centre of mass method incorporating background subtraction for accurate determination of fluorescence lifetime. *Optics Express*, 24(7):6899–6915, 2016. doi: 10.1364/OE.24.006899.

Polyakov, S. V. and Migdall, A. L. High accuracy verification of a correlated-photon-based method for determining photoncounting detection efficiency. *Optics Express*, 15(4):1390–1407, feb 2007. doi: 10.1364/OE.15.001390.

Popleteeva, M., Haas, K. T., Stoppa, D., Pancheri, L., Gasparini, L., Kaminski, C. F., Cassidy, L. D., Venkitaraman, A. R., and Esposito, A. Fast and simple spectral FLIM for biochemical and medical imaging. *Optics Express*, 23(18):23511–23525, sep 2015. doi: 10.1364/OE.23.023511.

Prabhu, K. M. M. *Window Functions and Their Applications in Signal Processing*. CRC Press, 2014. ISBN 978-1-4665-1584-0.

Preus, S. DecayFit, 2014. URL `www.fluortools.com`.

Puszka, A., Di Sieno, L., Dalla Mora, A., Pifferi, A., Contini, D., Boso, G., Tosi, A., Hervé, L., Planat-Chrétien, A., Koenig, A., and Dinten, J.-M. Time-resolved diffuse optical tomography using fast-gated single-photon avalanche diodes. *Biomedical Optics Express*, 4(8):1351–1365, aug 2013. doi: 10.1364/BOE.4.001351.

Raele, M. P., Tarelho, L. V. G., da Silva Azeredo, C. L., Couceiro, I. B., and Freitas, A. Z. Improving axial resolution in spectral domain low-coherence interferometry through fast Fourier transform harmonic artifacts. *Optical Engineering*, 53(7):073106, jul 2014. doi: 10.1117/1.OE.53.7.073106.

Rech, I., Luo, G., Ghioni, M., Yang, H., Xie, X. S., and Cova, S. Photon-Timing Detector Module for Single-Molecule Spectroscopy With 60-ps Resolution. *IEEE Journal on Selected Topics in Quantum Electronics*, 10(4):788–795, 2004. doi: 10.1109/JSTQE.2004.833975.

Rech, I., Ingargiola, A., Spinelli, R., Labanca, I., Marangoni, S., Ghioni, M., and Cova, S. Optical crosstalk in single photon avalanche diode arrays: a new complete model. *Optics Express*, 16(12):8381–8394, jun 2008. doi: 10.1364/OE.16.008381.

Richardson, J., Walker, R., Grant, L., Stoppa, D., Borghetti, F., Charbon, E., Gersbach, M., and Henderson, R. K. A 32×32 50ps resolution 10 bit time to digital converter array in 130nm CMOS for time correlated imaging. In *Proceedings of the Custom Integrated Circuits Conference*, pages 77–80, 2009. ISBN 9781424440726. doi: 10.1109/CICC.2009.5280890.

Richardson, J. A., Webster, E. A. G., Grant, L. A., and Henderson, R. K. Scaleable Single-Photon Avalanche Diode Structures in Nanometer CMOS Technology. *IEEE Transactions on Electron Devices*, 58(7):2028–2035, jul 2011. doi: 10.1109/TED.2011.2141138.

Ritter, J. W. Auffindung nicht sichtbarer Sonnenstrahlen außerhalb des Farbenspectrums, an der Seite des Viotets. *Annalen der Physik*, 7:501–528, 1801.

Schmid, T., Opilik, L., Blum, C., and Zenobi, R. Nanoscale Chemical Imaging Using Tip-Enhanced Raman Spectroscopy: A Critical Review. *Angewandte Chemie - International Edition*, 52(23):5940–5954, jun 2013. doi: 10.1002/anie.201203849.

Sharma, U., Chang, E. W., and Yun, S. H. Long-wavelength optical coherence tomography at 1.7 microm for enhanced imaging depth. *Optics Express*, 16(24):19712–19723, nov 2008. doi: 10.1364/oe.16.019712.

Sharman, K. K., Periasamy, A., Ashworth, H., Demas, J. N., and Snow, N. H. Error analysis of the rapid lifetime determination method for double-exponential decays and new windowing schemes. *Analytical Chemistry*, 71(5):947–952, 1999. doi: 10.1021/ac981050d.

Sheppard, C. J. Depth of field in optical microscopy. *Journal of Microscopy*, 149(1):73–75, 1988. doi: 10.1111/j.1365-2818.1988.tb04563.x.

Shreve, A. P., Cherepy, N. J., and Mathies, R. A. Effective Rejection of Fluorescence Interference in Raman Spectroscopy Using a Shifted Excitation Difference Technique. *Applied Spectroscopy*, 46(4):707–711, apr 2005. doi: 10.1366/0003702924125122.

Simms, W. On the Optical Glass prepared by the late Dr. Ritchie. *Memoirs of the Astronomical Society of London*, 11:165–170, 1840.

Smith, W. Effect of Light on Selenium During .the Passage of An Electric Current*. *Nature*, 7:303, 1873. doi: 10.1038/007303e0.

Sorin, W. V. and Baney, D. M. A Simple Intensity Noise Reduction Technique for Optical Low-Coherence Reflectometry. *IEEE Photonics Technology Letters*, 4(12): 1404–1406, 1992. doi: 10.1109/68.180591.

Spinelli, A. and Lacaita, A. L. Physics and Numerical Simulation of Single Photon Avalanche Diodes. *IEEE Transactions on Electron Devices*, 44(11):1931–1943, 1997. doi: 10.1109/16.641363.

Standish, B. A., Lee, K. K., Mariampillai, A., Munce, N. R., Leung, M. K., Yang, V. X., and Vitkin, I. A. In vivo endoscopic multi-beam optical coherence tomography. *Physics in Medicine and Biology*, 55(3):615–622, jan 2010. doi: 10.1088/0031-9155/55/3/004.

Stiles, P. L., Dieringer, J. A., Shah, N. C., and Van Duyne, R. P. Surface-enhanced Raman spectroscopy. *Annual Review of Analytical Chemistry*, 1:601–626, jul 2008. doi: 10.1146/annurev.anchem.1.031207.112814.

Stipčević, M. Active quenching circuit for single-photon detection with Geiger mode avalanche photodiodes. *Applied Optics*, 48(9):1705–1714, 2009. doi: 10.1364/AO.48.001705.

Stirbet, A. and Govindjee. On the relation between the Kautsky effect (chlorophyll a fluorescence induction) and Photosystem II: Basics and applications of the OJIP fluorescence transient. *Journal of Photochemistry and Photobiology B: Biology*, 104 (1-2):236–257, 2011. doi: 10.1016/j.jphotobiol.2010.12.010.

Suhling, K., Hirvonen, L. M., Levitt, J. A., Chung, P.-H., Tregidgo, C., Le Marois, A., Rusakov, D. A., Zheng, K., Ameer-Beg, S., Poland, S., Coelho, S., Henderson, R., and Krstajic, N. Fluorescence lifetime imaging (FLIM): Basic concepts and some recent developments. *Medical Photonics*, 27:3–40, may 2015. doi: 10.1016/j.medpho.2014.12.001.

Sun, Y., Liu, R., Elson, D. S., Hollars, C. W., Jo, J. A., Park, J., Sun, Y., and Marcu, L. Simultaneous time- and wavelength-resolved fluorescence spectroscopy for near real-time tissue diagnosis. *Optics Letters*, 33(6):630–632, mar 2008. doi: 10.1364/OL.33.000630.

Sun, Y., Sun, Y., Stephens, D., Xie, H., Phipps, J., Saroufeem, R., Southard, J., Elson, D. S., and Marcu, L. Dynamic tissue analysis using time- and wavelength-resolved fluorescence spectroscopy for atherosclerosis diagnosis. *Optics Express*, 19(5): 3890–3901, feb 2011. doi: 10.1364/OE.19.003890.

Swanson, E. A., Huang, D., Fujimoto, J. G., Hee, M. R., Lin, C. P., and Puliafito, C. A. High-speed optical coherence domain reflectometry. *Optics Letters*, 17(2):151–153, jan 1992. doi: 10.1364/ol.17.000151.

Swanson, E. A. and Fujimoto, J. G. The ecosystem that powered the translation of OCT from fundamental research to clinical and commercial impact [Invited]. *Biomedical Optics Express*, 8(3):1638–1664, mar 2017. doi: 10.1364/boe.8.001638.

Szabo, T. L. *Diagnostic Ultrasound Imaging : Inside Out*, volume 25. Academic Press, 2nd edition, 2014. ISBN 978-0-12-396487-8. doi: 10.1016/C2011-0-07261-7.

Takai, I., Matsubara, H., Soga, M., Ohta, M., Ogawa, M., and Yamashita, T. Single-Photon avalanche diode with enhanced NIR-sensitivity for automotive LIDAR systems. *Sensors*, 16(4):459, mar 2016. doi: 10.3390/s16040459.

Talbot, W. H. F. Some Experiments on Coloured Flames. *Edinburgh Journal of Science*, 5:77–81, 1826.

Tellinghuisen, J. and Wilkerson Jr., C. W. Bias and Precision in the Estimation of Exponential Decay Parameters from Sparse Data. *Analytical Chemistry*, 65(9): 1240–1246, may 1993. doi: 10.1021/ac00057a022.

Thomas, N. C. The Early History of Spectroscopy. *Journal of Chemical Education*, 68 (8):631–634, 1991. doi: 10.1021/ed068p631.

Thomson, J. J. Ionisation by Light. Photo-Electic Effects. In *Conduction of Electricity Through Gases*, chapter X., pages 250–290. Cambridge University Press, 2nd edition, 1906.

Tolles, W. M., Nibler, J. W., McDonald, J. R., and Harvey, A. B. A Review of the Theory and Application of Coherent Anti-Stokes Raman Spectroscopy (CARS). *Applied Spectroscopy*, 31(4):253–271, jul 1977. doi: 10.1366/000370277774463625.

Tomlins, P. H., Smith, G. N., Woolliams, P. D., Rasakanthan, J., and Sugden, K. Femtosecond laser micro-inscription of optical coherence tomography resolution test artifacts. *Biomedical Optics Express*, 2(5):1319–1327, may 2011. doi: 10.1364/boe.2.001319.

Valeur, B. and Brochon, J.-C., editors. *New Trends in Fluorescence Spectroscopy*, volume 1 of *Springer Series on Fluorescence*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001. ISBN 978-3-642-63214-3 (print) 978-3-642-56853-4 (eBook). doi: 10.1007/978-3-642-56853-4.

van den Bos, A. and Swarte, J. H. Resolvability of the Parameters of Multiexponentials And Other Sum Models. *IEEE Transactions on Signal Processing*, 41(1):313–322, jan 1993. doi: 10.1109/TSP.1993.193148.

Veerappan, C. and Charbon, E. CMOS SPAD Based on Photo-Carrier Diffusion Achieving PDP >40% from 440 to 580 nm at 4 v Excess Bias. *IEEE Photonics Technology Letters*, 27(23):2445–2448, 2015. doi: 10.1109/LPT.2015.2468067.

Veerappan, C., Richardson, J., Walker, R., Li, D.-U., Fishburn, M. W., Maruyama, Y., Stoppa, D., Borghetti, F., Gersbach, M., Henderson, R. K., and Charbon, E. A 160x128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter. In *2011 IEEE International Solid-State Circuits Conference*, pages 312–313, San Frasisco, CA, feb 2011. IEEE. ISBN 978-1-61284-302-5 (electronic), 978-1-61284-303-2 (print). doi: 10.1109/ISSCC.2011.5746333.

Villa, F., Markovic, B., Bellisai, S., Bronzi, D., Tosi, A., Zappa, F., Tisa, S., Durini, D., Weyers, S., Paschen, U., and Brockherde, W. SPAD Smart Pixel for Time-of-Flight and Time-Correlated Single-Photon Counting Measurements. *IEEE Photonics Journal*, 4(3):795–804, 2012. doi: 10.1109/JPHOT.2012.2198459.

Waeny, M. High-Speed Line-Scan Sensors Improve Spectroscopy, 2014. URL `https://www.photonics.com/Articles/High-Speed{_}Line-Scan{_}Sensors{_}Improve{_}Spectroscopy/a56993`. Accessed: 2019-02-06.

Ware, M., Migdall, A., Bienfang, J. C., and Polyakov, S. V. Calibrating photon-counting detectors to high accuracy: background and deadtime issues. In *Journal of Modern Optics*, volume 54, pages 361–372. Taylor & Francis Group, jan 2007. doi: 10.1080/09500340600759597.

Webster, E. A. G., Grant, L. A., and Henderson, R. K. A High-Performance Single-Photon Avalanche Diode in 130-nm CMOS Imaging Technology. *IEEE Electron Device Letters*, 33(11):1589–1591, 2012a. doi: 10.1109/LED.2012.2214760.

Webster, E. A. G., Richardson, J. A., Grant, L. A., Renshaw, D., and Henderson, R. K. A Single-Photon Avalanche Diode in 90-nm CMOS Imaging Technology With 44% Photon Detection Efficiency at 690 nm. *IEEE Electron Device Letters*, 33(5):694–696, 2012b. doi: 10.1109/LED.2012.2187420.

Weeks, M. E. The discovery of the elements. XIII. Some spectroscopic discoveries. *Journal of Chemical Education*, 9(8):1413–1434, 1932. doi: 10.1021/ed009p1413.

Wei, D., Chen, S., and Liu, Q. Review of Fluorescence Suppression Techniques in Raman Spectroscopy. *Applied Spectroscopy Reviews*, 50(5):387–406, 2015. doi: 10.1080/05704928.2014.999936.

Wheatstone, C. On the Prismatic Decomposition of Electrical Light. In *The Scientific Papers of Sir Charles Wheatstone*, pages 223–224. Cambridge University Press, 2011. ISBN 9781139057950 (online) 978-1-108-03202-5 (print). doi: 10.1017/CBO9781139057950.

William R. Leo. *Techniques for Nuclear and Particle Physics Experiments*. Springer-Verlag Berlin Heidelberg, 2nd edition, 1994. ISBN 978-3-642-57920-2 (eBook) 978-3-540-57280-0 (print). doi: 10.1007/978-3-642-57920-2.

Wojtkowski, M., Kowalczyk, A., Leitgeb, R., and Fercher, A. F. Full range complex spectral optical coherence tomography technique in eye imaging. *Optics Letters*, 27 (16):1415–1417, aug 2002. doi: 10.1364/OL.27.001415.

Wojtkowski, M., Srinivasan, V. J., Ko, T. H., Fujimoto, J. G., Kowalczyk, A., and Duker, J. S. Ultrahigh-resolution, high-speed, Fourier domain optical coherence tomography and methods for dispersion compensation. *Optics Express*, 12(11): 2404–2422, may 2004. doi: 10.1364/opex.12.002404.

Wollaston, W. H. A Method of Examining Refractive and Dispersive Powers, by Prismatic Reflection. *Philosophical Transactions of the Royal Society of London*, 92(0):365–380, 1802. doi: 10.1098/rstl.1802.0014.

Woolliams, P. D. and Tomlins, P. H. Estimating the resolution of a commercial optical coherence tomography system with limited spatial sampling. In *Measurement Science and Technology*, volume 22. IOP Publishing, 2011. doi: 10.1088/0957-0233/22/6/065502.

Woolliams, P. D., Ferguson, R. A., Hart, C., Grimwood, A., and Tomlins, P. H. Spatially deconvolved optical coherence tomography. *Applied Optics*, 49(11):2014–2021, apr 2010. doi: 10.1364/ao.49.002014.

Wu, C.-T., Tsai, M.-T., and Lee, C.-K. Two-Level Optical Coherence Tomography Scheme for Suppressing Spectral Saturation Artifacts. *Sensors*, 14(8):13548–13555, jul 2014. doi: 10.3390/s140813548.

Xie, H., Bec, J., Liu, J., Sun, Y., Lam, M., Yankelevich, D. R., and Marcu, L. Multispectral scanning time-resolved fluorescence spectroscopy (TRFS) technique for intravascular diagnosis. *Biomedical Optics Express*, 3(7):1521–1533, jul 2012. doi: 10.1364/BOE.3.001521.

Xu, H., Pancheri, L., Braga, L. H., Betta, G.-F. D., and Stoppa, D. Crosstalk characterization of single-photon avalanche diode (SPAD) arrays in CMOS 150nm technology. In *Procedia Engineering*, volume 87, pages 1270–1273. Elsevier, jan 2014. doi: 10.1016/j.proeng.2014.11.417.

Xu, J. and Knutson, J. R. Ultrafast Fluorescence Spectroscopy via Upconversion. Applications to Biophysics. *Methods in Enzymology*, 450:159–183, jan 2008. doi: 10.1016/S0076-6879(08)03408-3.

Xu, Y., Xiang, P., and Xie, X. Comprehensive understanding of dark count mechanisms of single-photon avalanche diodes fabricated in deep sub-micron CMOS technologies. *Solid-State Electronics*, 129:168–174, mar 2017. doi: 10.1016/j.sse.2016.11.009.

Yankelevich, D. R., Ma, D., Liu, J., Sun, Y., Sun, Y., Bec, J., Elson, D. S., and Marcu, L. Design and evaluation of a device for fast multispectral time-resolved fluorescence spectroscopy and imaging. *Review of Scientific Instruments*, 85:034303, mar 2014. doi: 10.1063/1.4869037.

Yaqoob, Z., Wu, J., McDowell, E. J., Heng, X., and Yang, C. Methods and application areas of endoscopic optical coherence tomography. *Journal of Biomedical Optics*, 11(6):063001, 2006. doi: 10.1117/1.2400214.

Yaseen, M. A., Sakadžić, S., Wu, W., Becker, W., Kasischke, K. A., and Boas, D. A. In vivo imaging of cerebral energy metabolism with two-photon fluorescence lifetime microscopy of NADH. *Biomedical Optics Express*, 4(2):307, 2013. doi: 10.1364/boe.4.000307.

Yun, S., Tearney, G. J., Bouma, B. E., Park, B. H., and de Boer, J. F. High-speed spectral-domain optical coherence tomography at 1.3 mum wavelength. *Optics Express*, 11(26):3598–3604, dec 2003. doi: 10.1364/OE.11.003598.

Yun, S. H., Tearney, G. J., de Boer, J. F., and Bouma, B. E. Motion artifacts in optical coherence tomography with frequency-domain ranging. *Optics Express*, 12(13):2977, jun 2004. doi: 10.1364/opex.12.002977.

Zhang, J., Itzler, M. A., Zbinden, H., and Pan, J. W. Advances in InGaAs/InP single-photon detector systems for quantum communication. *Light: Science and Applications*, 4(5):e286, may 2015. doi: 10.1038/lsa.2015.59.

Zhou, C., Wang, J., and Jiao, S. Dual channel dual focus optical coherence tomography for imaging accommodation of the eye. *Optics Express*, 17(11):8947–8955, may 2009. doi: 10.1364/OE.17.008947.

Ziarkash, A. W., Joshi, S. K., Stipčević, M., and Ursin, R. Comparative study of afterpulsing behavior and models in single photon counting avalanche photo diode detectors. *Scientific Reports*, 8:5076, dec 2018. doi: 10.1038/s41598-018-23398-z.

Zworykin, V. K., Morton, G. A., and Malter, L. The Secondary Emission Multiplier-A New Electronic Device. *Proceedings of the Institute of Radio Engineers*, 24(3):351–375, mar 1936. doi: 10.1109/JRPROC.1936.226435.